

RITA MARIA ALVES DE MORAES

INTROGRESSÃO DE ALELOS PARA ALTO TEOR DE PROTEÍNA EM
SOJA ASSISTIDA POR MARCADORES MOLECULARES

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de “Doctor Scientiae”.

VIÇOSA
MINAS GERAIS – BRASIL
2003

RITA MARIA ALVES DE MORAES

INTROGRESSÃO DE ALELOS PARA ALTO TEOR DE PROTEÍNA EM
SOJA ASSISTIDA POR MARCADORES MOLECULARES

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de "Doctor Scientiae".

APROVADA: 04 de abril de 2003

Prof. Everaldo Gonçalves de Barros
(Conselheiro)

Prof. Cosme Damião Cruz

Prof. Valéria Monteze Guimarães

Dr. Ivan Schuster

Prof. Maurilio Alves Moreira
(Orientador)

À minha mãe, Maria de Lourdes.

À minha vovó Luzia (*in memoriam*).

Aos meus amigos.

Dedico.

AGRADECIMENTOS

A Deus, por tudo que existe.

À Universidade Federal de Viçosa (UFV), pela oportunidade de realizar o curso.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de estudos.

Ao Instituto de Biotecnologia Aplicada à Agropecuária (BIOAGRO), pela infra-estrutura adequada ao desenvolvimento de trabalhos científicos.

Ao meu estimado orientador professor Maurilio Alves Moreira, pelo exemplo de profissional, pelo carinho e pelo apoio sempre presentes.

Ao professor Everaldo Gonçalves de Barros, pela orientação, pela amizade e pelo carinho.

Ao professor Cosme Damião Cruz, pela orientação e por sempre ter tempo para os seus alunos.

Aos mestres, que fizeram do magistério um ideal, mesclando a arte de ensinar com o dom da convivência, tornando-se amigos, transmitindo suas experiências que enormemente ajudaram em minha formação. Meu perdão àqueles que se limitaram em ser apenas professores.

Ao Newton, por suas sugestões e pela amizade.

À Inês, por sua ajuda, pelo apoio e pela amizade.

Ao Ivan Schuster, por seu apoio e pela sua amizade.

Aos meus amigos do laboratório de Proteína: Lucinete, Telma, Taís, Valéria, Simone, e Antônio Vieira, onde está minha segunda casa.

Aos estudantes de iniciação científica: Josie, Fernanda, Yaska, e Gustavo que sempre me fazem aprender coisas novas.

Ao João Paulo (meu braço direito), Naldo e Jander, os quais tanto colaboraram na elaboração deste trabalho.

Aos colegas do Biomol: Fábio, Vilmar, Gerardo, Lucimara, Márcia (Flores), Márcia, Fernanda, Luciano, Pedro, Klever, Demerson, Thiago, Marcelo, pelo agradável convívio.

Aos funcionários do BIOAGRO: Aloísio, Domingos, Gláucia, Expedito, Fausto, Eugênio, Edson, Alessandra, Luzia, Élcio e Márcio pelo convívio.

A todos os colegas dos demais laboratórios pelo apoio e pelo convívio.

À tia Rose, pelo incentivo.

Aos funcionários, José Carlos, Marcos e Cupertino pela colaboração na casa de vegetação.

Às secretárias, Rita de Cássia e Conceição, pela ajuda e dedicação durante esses seis anos de curso

À Sueli, por manter minha casa como um lar e pela sua amizade.

A todos aqueles que colaboraram e torceram pelo meu sucesso.

BIOGRAFIA

RITA MARIA ALVES DE MORAES, filha de Maria de Lourdes Alves de Moraes e José Batista de Moraes, nasceu em Itabuna, Bahia, onde cursou o primeiro e segundo graus.

Em março de 1989, ingressou no Curso de Agronomia da Universidade Federal de Viçosa (UFV), em Viçosa, MG, diplomando-se em agosto de 1994.

Foi bolsista de aperfeiçoamento de fevereiro de 1996 a fevereiro de 1997.

Em março de 1997, iniciou o curso de Mestrado em Genética e Melhoramento na UFV, defendendo a tese em 15 de março de 1999.

Em abril de 1999, iniciou o curso de Doutorado em Genética e Melhoramento na UFV, defendendo a tese em 04 de abril de 2003.

CONTEÚDO

RESUMO.....	ix
ABSTRACT.....	xi
1. INTRODUÇÃO GERAL.....	1
2. OBJETIVOS.....	4
3. REVISÃO BIBLIOGRÁFICA.....	5
3.1. Origem e cultura da soja no Brasil.....	5
3.2. O melhoramento da soja.....	6
3.3. Proteína de soja.....	7
3.3.1. Herdabilidade do teor de proteína em soja.....	12
3.3.2. Melhoramento para teor de proteína em soja.....	13
3.4. Melhoramento genético de plantas auxiliado por marcadores moleculares.....	18
4.REFERÊNCIAS BIBLIOGRÁFICAS.....	23
CAPÍTULO 1	
DIVERGÊNCIA GENÉTICA ENTRE PROGENITORES PARA USO EM PROGRAMAS DE RETROCRUZAMENTOS.....	31
RESUMO.....	32
INTRODUÇÃO.....	33
MATERIAL E MÉTODOS.....	35
RESULTADOS E DISCUSSÃO.....	39
REFERÊNCIAS BIBLIOGRÁFICAS.....	49

CAPÍTULO 2	
ESTIMAÇÃO DO NÚMERO REFERENCIAL DE <i>PRIMERS</i> DE MICROSSATÉLITES POR SIMULAÇÃO PARA ASSISTIR UM PROGRAMA DE RETROCRUZAMENTOS EM SOJA.....	51
RESUMO.....	52
INTRODUÇÃO.....	54
MATERIAL E MÉTODOS.....	58
RESULTADOS E DISCUSSÃO.....	64
REFERÊNCIAS BIBLIOGRÁFICAS.....	73
ANEXO DO CAPÍTULO 2.....	75
CAPÍTULO 3	
INTROGRESSÃO DE ALELOS PARA ALTO TEOR DE PROTEÍNA EM VARIEDADES DE SOJA.....	93
RESUMO.....	94
INTRODUÇÃO.....	96
MATERIAL E MÉTODOS.....	99
RESULTADOS E DISCUSSÃO.....	104
REFERÊNCIAS BIBLIOGRÁFICAS.....	122
CAPÍTULO 4	
CARACTERIZAÇÃO BIOQUÍMICA DE LINHAGENS DE SOJA COM ALTO TEOR DE PROTEÍNA.....	123
RESUMO.....	124
INTRODUÇÃO.....	126
MATERIAL E MÉTODOS.....	128
RESULTADOS E DISCUSSÃO.....	132

REFERÊNCIAS BIBLIOGRÁFICAS.....	138
---------------------------------	-----

RESUMO

MORAES, Rita Maria Alves de, D.S., Universidade Federal de Viçosa, Abril de 2003. **Introgressão de alelos para alto teor de proteína em soja assistida por marcadores moleculares.** Orientador: Maurilio Alves Moreira. Conselheiros: Everaldo Gonçalves de Barros e Sergio H. Brommonschenkel.

O teor de proteína no farelo de soja tem influenciado o preço final pago ao produtor. Por esta razão, um dos principais objetivos do melhoramento da soja é o desenvolvimento de variedades com alto teor de proteína e alta produtividade. Três variedades comerciais e seis linhagens de soja com teor normal de proteína foram cruzadas com dezoito linhagens de alto teor de proteína para promover uma combinação desejável de alto teor de proteína e alta produtividade. Os objetivos gerais deste trabalho foram a introgressão de alelos de alto teor de proteína em variedades elite de soja por retrocruzamentos e seleção assistida por microanálises bioquímicas e por marcadores microssatélites. Neste sentido, várias metas específicas foram atingidas neste trabalho tais como: (a) realização de uma análise de divergência genética baseado em marcadores microssatélites entre progenitores recorrentes e doadores de alto teor de proteína; (b) determinação do número ideal de *primers* de microssatélites para assistir um programa de retrocruzamentos; (c) realização de seleção por meio de microanálises bioquímicas após uma única geração de autofecundação; (d) seleção de indivíduos mais próximos geneticamente do progenitor recorrente e com maior teor de proteína; (e) determinação do ganho de seleção das plantas selecionadas e (f) caracterização de linhagens de soja com alto teor de proteína quanto à sua composição bioquímica. Como conclusões, foi evidenciado que a utilização de 57 pares de *primers* de microssatélites (SSR) foram eficientes para avaliar a diversidade genética entre os genótipos doadores e recorrentes, permitindo que genótipos mais próximos

geneticamente entre si fossem utilizados no programa de retrocruzamento para introgressão de alelos de alto teor de proteína em soja. Foram definidos 44 dos 57 pares de *primers* de SSR previamente selecionados como o número ótimo para assistir o programa de retrocruzamentos visando a introgressão de alelos de alto teor de proteína, com 95% de correlação e 6,44% de estresse e, ainda, foram escolhidas três combinações destes 44 pares de *primers* que proporcionaram o mesmo agrupamento original obtido quando foram utilizados 57 pares de *primers*. A análise de variância evidenciou variabilidade genética para todos os cruzamentos avaliados e um alto valor de herdabilidade para teor de proteína em soja, sendo que os melhores cruzamentos foram aqueles que envolveram o progenitor recorrente OC 953312, que apresentou a maior herdabilidade (93,1%) e maior variabilidade genética (4,62). Os melhores progenitores doadores foram escolhidos com base em suas capacidade específica de combinação (CEC), por meio do teste de média, o que reduzirá o número de cruzamentos e análises e, conseqüentemente, os custos do programa de retrocruzamento. As plantas selecionadas para alto teor de proteína independente do cruzamento tiveram em média teores de proteína de 45,74 a 47,47% sendo que os ganhos reais para teor de proteína foram de 3,88 a 9,35%. As plantas que tiveram alto teor de proteína e menores distâncias genéticas em relação ao progenitor recorrente foram selecionadas para dar prosseguimento ao programa de retrocruzamentos. A caracterização bioquímica das isolinhas com alto teor de proteína indicou que o aumento no teor de proteína promoveu uma redução no teor de óleo e carboidratos totais. A análise da composição aminoacídica das isolinhas de alto teor de proteína comparada com a linhagem recorrente (normal) evidenciou que não houve alteração aparente na composição aminoacídica das isolinhas e, ainda, foi observado que o aumento do teor de proteína ocorreu principalmente devido ao aumento no teor da proteína de reserva 11S (de melhor qualidade), havendo inclusive uma pequena redução na proteína de reserva 7S. Esta observação é importante, pois indica que o melhoramento para teor de proteína não reduziu o valor nutricional da proteína da soja, podendo até mesmo melhorar a sua qualidade.

ABSTRACT

MORAES, Rita Maria Alves de, D.S., Universidade Federal de Viçosa, April, 2003. **Introgression of alleles for high protein content into soybean elite varieties assisted by molecular markers.** Advisor: Maurilio Alves Moreira. Committee members: Everaldo Gonçalves de Barros and Sergio H. Brommonschenkel.

Soy flour protein content greatly influences the final price paid to the producer. For this reason one of the main objectives of soybean breeding is to select for high protein cultivars. The main goal of this work was the introgression of alleles for high protein content in soybean elite cultivars through backcross assisted by molecular markers. For that reason, several specific tasks were achieved, such as (a) analyses of genetic diversity among recurrent and donor parents with microsatellite markers; (b) determination of the ideal number of microsatellite primers to assist the backcross program; (c) selection of high protein genotypes by using non-destructive single seed analyses after one generation of self-pollination; (d) selection of genotypes genetically closer to the recurrent parent; (e) protein yield determination of F2 selected plants and (f) biochemical characterization of high protein soybean lines. The use of 57 selected pairs of microsatellite primers was efficient to evaluate the genetic diversity among the recurrent and donor parents, which allow the choice of genetically closer genotypes for the backcross breeding program. Forty-four of the 57 pairs of primers previously selected were defined as an optimum number of primers to assist the backcross breeding program, with 95% correlation and 6.44% stress. The three best 44 primers combinations that yields the same original clustering obtained with the 57 pairs of *primers* were chosen. Variance analysis evidenced genetic variability for all crosses analyzed and a high heritability value for protein content in soybean, being the best crosses the ones that used recurrent parent OC 953312, which showed the highest heritability

(93.1%) and genetic variability values (4.62). The best donor parents were chosen based on capacity specific combination by using the mean test. This will reduce the number of crosses and biochemical analyses needed and, consequently, the cost of the backcross breeding program. Selected plants for high protein independently of the crossing had an average of protein content varying from 45.74 to 47.47%, and the real gains for protein varying from 3.88 (9.17%) to 9.35 (24.86%). Plants that had the highest protein content and the lowest genetic distances in relation to the recurrent parent were selected for the backcross breeding. A biochemical characterization was performed in the high protein near isogenic lines (NILs) derived from CAC 1 variety. The increase in protein content promotes reduction in oil as well as in total carbohydrates contents of the soybean seeds. A soy flour amino acid analysis showed that there was no significant alteration in the amino acid contents in the high protein lines. Scanner densitometry estimation of soybean protein components separated by SDS-PAGE showed that the increase of protein content of the NILs are due to an increase in the 11S storage protein component, which is of better quality. The scanner densitometry evaluation showed a slight reduction in 7S storage protein content in the high protein lines. This observation is of interest since it pointed out that the backcross breeding for high protein content does not reduce, being even possible to increase the nutritional and functional values of the soybean protein.

1. INTRODUÇÃO GERAL

Os principais objetivos do melhoramento de plantas, ao longo dos anos, têm sido aumentar a produção e melhorar a qualidade dos grãos produzidos de forma a suprir a população com uma fonte adequada de alimentos. Com o surgimento de novas metodologias genéticas, foi possível promover a geração e a seleção de combinações desejáveis de genes. Embora as novas técnicas de melhoramento genético aumentassem a probabilidade de se encontrar combinações úteis desejáveis, estas são dependentes de grandes populações de plantas. Com o advento das técnicas de biologia molecular, tem sido possível encontrar combinações específicas de genes desejáveis provenientes de diversas fontes, eliminar genes indesejáveis e utilizar a engenharia genética com a finalidade de obter características antes não encontradas na espécie. Portanto, a aplicação das técnicas de biologia molecular, como ferramenta auxiliar, vem corroborar com o melhoramento convencional, promovendo redução no tempo e no tamanho das populações requeridas para a recuperação de germoplasma elite.

O programa de melhoramento genético da qualidade da soja em desenvolvimento no BIOAGRO/Universidade Federal de Viçosa foi iniciado com o principal objetivo de criar cultivares de soja mais adequados para a agroindústria. Dentro deste programa, dois caracteres considerados de grande importância para aumentar a utilização da soja pela indústria de alimentos foram, inicialmente, incorporados, ou seja, (1) redução significativa do “beany flavor”, determinado pela presença das enzimas lipoxigenases, o qual é responsável pela baixa aceitação e alta instabilidade dos produtos derivados da soja e (2) presença de tegumento e hilo claros, caracteres determinantes da qualidade e aceitabilidade da adição de produtos protéicos de soja em alimentos industrializados. Além desses, outros caracteres têm sido incorporados no programa de melhoramento da qualidade da soja do BIOAGRO/UFV, como (3) ausência do inibidor de tripsina Kunitz e de (4) lectina (ambos são considerados fatores antinutricionais), (5) altos níveis de proteína (algumas linhagens têm apresentado 52% de proteína),

modificações na fração óleo como (6) baixos níveis de ácido linolênico (algumas linhagens têm apresentado 2 a 3%, enquanto as normais apresentam de 8 a 12%) e (7) altos níveis de ácido oléico (tais modificações são consideradas desejáveis pois promovem maior estabilidade e melhor qualidade do óleo de soja para fins industriais). Outros caracteres desejáveis como redução genética dos teores dos açúcares rafinose e derivados (RO) e aumento dos teores de isoflavonas estão, no momento, sendo avaliados. A maioria destas modificações genéticas tem sido introduzida e piramidada em cultivares elites por meio de retrocruzamentos assistidos por microanálises bioquímicas não destrutivas e por marcadores moleculares de DNA.

Como parte deste programa de qualidade da soja para a agroindústria, o melhoramento da soja para alto teor de proteína, será enfatizado neste trabalho.

O farelo de soja destinado à exportação é classificado, de acordo com o seu conteúdo de proteína, nas categorias “HyPro” (> 48%), “Normal” (46%) e “LowPro” (< 43,5%). Lotes classificados como HyPro recebem ágio sobre o valor normal de comercialização, enquanto que lotes classificados com “LowPro” sofrem deságio sobre este valor. Para atingir o nível classificado como “Normal” e “HyPro” a soja deve conter, no mínimo, 41,5% e 43% de proteína nas sementes, respectivamente, com base na matéria seca (informações obtidas junto a Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda – COODETEC, Cascavel, PR).

A grande maioria das variedades brasileiras de soja, em especial aquelas cultivadas na região sul, possui teor de proteína considerado baixo. Isto porque, os programas de melhoramento, ao longo dos anos, selecionaram linhagens apenas em função da característica produtividade, a qual possui correlação negativa com o conteúdo de proteína dos grãos.

Trabalhos tradicionais de retrocruzamentos que têm como finalidade aumentar o teor de proteína em sementes de soja demandam longo tempo. A seleção geralmente é feita após 2-3 ciclos de autofecundação, em cada geração de retrocruzamento, devido à utilização de métodos destrutivos de análise. Neste particular, métodos de determinação de proteínas que não prejudicam a germinação posterior das sementes analisadas são muito úteis

na análise não destrutiva de sementes individuais, permitindo que sementes com alto teor de proteína sejam selecionadas após a primeira geração de autofecundação.

Em trabalhos de retrocruzamento, a seleção de indivíduos geneticamente mais próximos do progenitor recorrente reduz o número de retrocruzamentos necessários para a recuperação do genoma do progenitor recorrente. Marcadores moleculares são muito úteis em trabalhos de seleção, pois permitem avaliar regiões do genoma distribuídas em todos os cromossomos, resultando em um melhor acompanhamento da introgressão de genes durante o processo de retrocruzamento.

2. OBJETIVOS

Os principais objetivos deste trabalho foram:

2.1. Realizar análise de divergência genética entre progenitores (recorrentes e doadores) para alelos de alto teor de proteína visando orientar um programa de retrocruzamentos em soja;

2.2. Determinar o número de pares de *primers* de microssatélites necessários para acompanhar a recuperação do genoma do progenitor recorrente em cada ciclo de retrocruzamento em soja;

2.3. Realizar seleção para alto teor de proteína em soja após uma única geração de autofecundação, utilizando métodos não destrutivos de análise;

2.4. Selecionar genótipos de alto teor de proteína geneticamente mais próximos do progenitor recorrente, com auxílio de marcadores moleculares visando reduzir o número de retrocruzamentos em soja;

2.5. Determinar o ganho de seleção e a eficiência do uso de marcadores moleculares na introgressão de genes para alto teor de proteína em soja;

2.6. Caracterizar linhagens de soja com alto teor de proteína quanto à composição aminoacídica, relação 7S/11S, conteúdo de proteína, óleo e carboidratos totais.

3. REVISÃO BIBLIOGRÁFICA

3.1. Origem e cultura da soja no Brasil

O gênero *Glycine* é composto de dois subgêneros: *Glycine* e *Soja*. A soja perene selvagem pertence ao subgênero *Glycine* e tem vários arranjos genômicos. A soja cultivada (*Glycine max*) e seu ancestral anual selvagem (*Glycine soja*) pertencem ao subgênero *Soja*, contêm número de cromossomos $2n = 40$, são compatíveis sexualmente, podendo produzir híbridos viáveis e carregam genomas similares (PALMER *et al.* 1996).

BURTON (1997) demonstra que a soja foi domesticada no século XI AC no nordeste da China e que para localizar o ponto provável de origem, a similaridade entre a soja *Glycine max* e o seu parental selvagem mais próximo *Glycine soja*, foi investigado em variedades (land races) por toda a China. Deste estudo, concluiu-se que as duas espécies eram muito próximas quanto ao teor de proteína, frequência dos alelos do inibidor de tripsina e data de florescimento na latitude 35°N, e que o vale do Rio Amarelo seria o local de origem da soja. Dali, ela se disseminou para o sul da China e para o leste através da Coréia e Japão.

Aproximadamente 80% do germoplasma presente nos cultivares atuais pode ser rastreado a doze linhagens, as quais foram introduzidas nos Estados Unidos da América no início do século XX (LORENZEN 1995).

HIROMOTO e VELLO (1986) determinaram a base genética do germoplasma da soja cultivada no Brasil e compararam o grau de similaridade do germoplasma brasileiro com o norte-americano. Utilizando o coeficiente de parentesco de Malécot, esses autores identificaram 26 ancestrais do terceiro ciclo de melhoramento da soja no Brasil, sendo que 11 progenitores contribuíram individualmente com proporções variáveis de 3 a 15% e contribuição acumulada de 89% do conjunto gênico das variedades brasileiras. Além disso, seis destes progenitores são também os mais freqüentes nas variedades norte-americanas. Alguns destes ancestrais foram procedentes das mesmas regiões da China e da Coréia. Esses dados mostram que a base genética do germoplasma de soja, adaptado a condições brasileiras, é muito restrita.

ABDELNOOR *et al.* (1995) usaram marcadores RAPD para avaliar a diversidade genética entre 38 variedades brasileiras de soja. Esta análise permitiu separar as variedades em cinco subgrupos distintos, e houve concordância dos resultados com o coeficiente de parentesco.

3.2. Melhoramento da soja

BURTON (1997) mostrou que o melhoramento da soja nos Estados Unidos tem sido visto como um programa de seleção recorrente onde cultivares são selecionados e liberados para produção, sendo recombinadas e então re-selecionadas. Este autor afirma ainda que a incorporação de genes para resistência a doenças e pragas tem servido para a proteção e manutenção do aumento genético na produção de grãos.

Atualmente um programa de melhoramento tradicional pode levar dez anos ou mais desde a escolha de progenitores até o estágio de produção do grão. O primeiro passo no melhoramento de plantas é a criação de uma população que seja geneticamente divergente pelo cruzamento entre progenitores contrastantes ou por mutação induzida no germoplasma melhorado.

PAULS (1995) relata que em um programa de melhoramento de plantas convencional a escolha dos progenitores é feita em um banco de germoplasma que contenha variedades registradas, variedades adaptadas da região ("land races"), germoplasma não melhorado coletado de material selvagem, espécies próximas, e que, atualmente, variedades geneticamente modificadas também estão sendo usadas no melhoramento. A escolha dos progenitores é uma decisão crucial para o melhorista tendo em vista que a diversidade genética que é incorporada no primeiro cruzamento determina as várias características que podem ser expressas nas variedades desenvolvidas, bem como o tempo que esta variedade leva para chegar aos objetivos do programa. O objetivo do programa é que irá ditar a fonte de germoplasma onde os progenitores serão selecionados. Normalmente, a seleção dos progenitores é feita com base em produtividade e alguma característica específica tais como resistência a doenças, a pragas, estresse

ambiental e características de qualidade da semente (BURTON 1997 e PAULS 1995). As melhores características dos pais são combinadas por meio de cruzamentos entre os mesmos. As progênies são avaliadas com base em vários critérios agrônômicos e as melhores plantas são usadas para novos cruzamentos e testes. As gerações são então avançadas para atingir a homozigose. A “estabilização genética” é necessária para aumentar a uniformidade e a estabilidade de produção ano após ano no campo. Após a variedade ser desenvolvida ela é testada em vários locais por várias gerações. Os resultados destes testes são usados para a liberação da variedade para plantio comercial.

3.3. Proteína de soja

Sementes de variedades cultivadas de soja contém cerca de 40% de proteína por peso de matéria seca. Muitas proteínas diferentes são encontradas na semente de soja, mas somente poucas estão presentes em grandes quantidades. As proteínas presentes em maior quantidade são: glicinina, β -conglucina, lipoxigenase, inibidor de tripsina Kunitz (KTI), inibidores de protease de baixo peso molecular (dos quais o melhor estudado é o inibidor Bowman-Birk) lectina e urease. Destas, glicinina e β -conglucina são as proteínas predominantes, perfazendo um total de cerca de 70% da proteína da semente (HILL & BREIDENBACH 1974). As outras, as quais estão presentes em altos níveis comparados com outras proteínas metabólicas perfazem, geralmente, cerca de 2 a 5% cada do total de proteína. As proteínas de reserva da soja têm coeficientes de sedimentação de, aproximadamente, 7S e 11S. O termo proteína de reserva é dado às proteínas que estão localizadas no cotilédone em partículas denominadas “corpos protéicos” os quais são preenchidos com proteína sem nenhuma função catalítica (KOSHIYAMA 1987).

Um sistema de classificação geral para caracterizar proteínas vegetais é baseado na solubilidade das proteínas em vários solventes (SMITH e GRIERSON 1982). Quatro tipos ou classes de proteína foram estabelecidos: albumina, globulina, prolamina e glutenina. Albuminas constituem um grupo relativamente simples de proteínas de 12 a 100 kD.

Elas são encontradas, provavelmente, em todos os tecidos vivos, são solúveis em água e podem ser clivadas por enzimas ou ácidos e são desnaturadas pelo calor. Proteínas do tipo albumina geralmente precipitam em soluções com altas concentrações de sulfato de amônio entre 70 a 100% de saturação. Proteínas do tipo globulina são relativamente mais complexas o que é indicado pela variação no peso molecular (18 a 360 kDa), sendo solúveis em soluções salinas (0,5 a 1,0 M de NaCl). Globulinas contêm altos níveis de glutamina, asparagina e arginina e baixas concentrações dos aminoácidos sulfurados metionina e cisteína. Prolaminas são proteínas de baixo peso molecular (16 a 60 kDa) solúveis em álcool 70%. Estas proteínas são ricas em glutamina e prolina, mas deficientes em arginina, histidina, triptofano e lisina. As proteínas do tipo gluteninas (30 a 50 kDa) são praticamente insolúveis em água, parcialmente solúveis em álcool ou ácido diluído e solúveis em solventes alcalinos (SMITH e GRIERSON 1982). Em soja, proteínas do tipo globulinas perfazem cerca de 90% da proteína da semente. A distinção de duas frações diferentes dentro das globulinas foi feita com base no coeficiente de sedimentação de 7S e 11S, as quais foram denominadas β -conglucina e glicina, respectivamente. Outras globulinas de soja são listadas na Tabela 1 (HILL e BREIDENBACH 1974; KOSHIYAMA 1983).

Tabela 1 – Teor e conteúdo de aminoácidos sulfurados das globulinas presentes em sementes de soja.

S _{w20}	Nomenclatura	(% da globulina total de soja)	Peso molecular (kDa)	Aminoácido	
				Metionina (%)	Cisteína (%)
2S	α-Conglicinina	12,7	18-33	0,5	4,8
7S	γ-Conglicinina	3,0	104	1,4	2,2
7S	β-Conglicinina	35,0	141-171	1,4	2,2
11S	Glicinina	41,8	317-360	1,5	3,0
15S	Poli-glicinina	7,5	Nd	Nd	Nd

Nd: não determinado

A composição de subunidades da glicinina varia entre cultivares de soja promovendo polimorfismos desta proteína em sementes de soja (KITAMURA *et al.* 1980). MOREIRA *et al.* (1979, 1981) mostraram que os aminoácidos sulfurados não eram distribuídos igualmente entre as várias subunidades de glicinina. Esta proteína é formada por seis polipeptídeos de caráter ácido (A1a, A1b, A2, A3, A4 e A5) e cinco de caráter básico (B1a, B1b, B2, B3 e B4) que se encontram associados na proteína nativa de modo específico por ligações bissulfídicas, formando as subunidades A1aB2, A1bB1b, A2B1a, A3B4 e A5A4B3. Atualmente, essas subunidades são denominadas de G1 (A1aB2), G2 (A1bB1b) G3 (A2B1a) e G4 (A3B4) e G5 (A5A4B3).

STASWICK e NIELSEN (1983) também demonstraram variação na composição de subunidades da glicinina da cultivar 'Raiden', a qual apresenta ausência da subunidade A5A4B3 e sugerem a possibilidade de manipulação genética da qualidade nutricional da proteína da semente de soja por alteração da composição de subunidades da glicinina. Não somente a qualidade da fração glicinina é importante, mas também a sua quantidade. HUGHES e MURPHY (1983) mostraram que a quantidade de glicinina é variável entre cultivares submetidos às mesmas condições ambientais, e que

isto poderia ser importante na qualidade de alimentos. MURPHY e RESURRECION (1984) mostraram que houve uma grande influência ambiental sobre o teor de glicinina (11S) do que sobre o teor de β -conglucina (7S), havendo também influência genética, porém em menor intensidade nas variedades estudadas.

UTUMI *et al.* (1998) estudaram o efeito da ausência de lipoxigenases e das subunidades α' (componente da beta-conglucina) e G4 (componente da glicinina) sobre a qualidade da proteína de soja. Estes autores observaram que a eliminação genética das subunidades α' e G4 não reduziu o teor de proteína total. Nas linhagens com ausência de lipoxigenases houve redução drástica na produção de hexanal, um dos principais produtos responsáveis pelos sabores característicos de produtos de soja. Foi observada, também, uma diminuição no teor de metionina em farinhas desengorduradas de linhagens com ausência da subunidade α' e um aumento do teor deste aminoácido em linhagens com ausência da subunidade G4.

OGAWA *et al.* (1989) e KITAMURA *et al.* (1981) analisaram linhagens com variações no teor da proteína 7S e concluíram que uma diminuição na síntese da proteína 7S pode ser compensada por um aumento na síntese e deposição da proteína 11S (que tem maior conteúdo de metionina), mantendo assim o teor de proteína total não alterado. Conseqüentemente, isto aumentaria o conteúdo de aminoácidos sulfurados da proteína de soja.

KITAMURA *et al.* (1984) mostraram que as ausências das subunidades A5A4B3 de glicinina e α' da β -conglucina são determinadas por alelos recessivos de dois genes que segregam independentemente no cruzamento entre o cultivar 'Raiden', com ausência da subunidade A5A4B3, com o cultivar 'Keburi', com ausência da subunidade α' . Estes alelos foram denominados cgy_4 e cgy_1 para ausência de A5A4B3 e α' , respectivamente. TAKAHASHI *et al.* (1996) denominaram o alelo recessivo, induzido por irradiação com raios gama, que determina a ausência da subunidade α de cgy_2 . Como pode ser notado nesta breve descrição, existem vários alelos recessivos de ocorrência natural, ou induzidos por radiações gama que

determinam a ausência de subunidades componentes das principais proteínas de reserva da soja.

KITAMURA (1995) caracterizou várias linhagens mutantes quanto à composição protéica das suas subunidades. As globulinas 7S (composta pelas subunidades α' , α e β) e 11S (composta pelo grupo 1 $A_{1a}B_2$, $A_{1b}B_{1b}$, A_2B_{1a} e as subunidades A_3B_4 e $A_5A_4B_3$) são as duas principais proteínas de reserva da soja. Um mutante obtido por irradiação (por raios gama) na variedade Karikei 434, possuía ausência da subunidade α' e continha baixos níveis das subunidades α e β . A linhagem mutante, apesar da acentuada redução no conteúdo de 7S, continha um teor de proteína normal, no entanto, o teor da subunidade 11S aumentou cerca de 15% em relação às variedades normais e nenhum efeito deletério nos processos fisiológicos do desenvolvimento da semente foi observado. Outros mutantes obtidos e identificados foram analisados por KITAMURA (1995), por exemplo: uma linhagem com ausência de todo o grupo 1, o qual tem sido mostrado estar sob controle genético de um único alelo recessivo; uma outra linhagem mutante com ausência de A_3B_4 , obtida de um acesso de *Glycine soja*, o qual também se mostrou estar sob o controle genético de um único alelo recessivo. Como *Glycine soja* é compatível sexualmente com *Glycine Max*, estas duas linhagens mutantes foram cruzadas de forma a se obter linhagens com vários perfis protéicos, ou seja, mutantes simples, duplo-mutantes e triplo-mutantes. Os mutantes duplo e triplo apresentaram uma redução acentuada no teor de 11S e nenhum aumento aparente no teor de 7S. O contrário foi observado com os mutantes baixo em 7S, os quais aumentaram a produção de 11S para compensar a redução nos níveis de 7S. Nenhum efeito deletério sobre processos fisiológicos foi observado, no entanto, se esta redução afeta negativamente o teor de proteína ou positivamente o teor de óleo em sementes de soja é uma questão que ainda precisa ser estudada.

3.3.1. Herdabilidade do teor de proteína em soja

A herdabilidade de um caráter métrico é uma das mais importantes propriedades de uma população, porque expressa a proporção da variância que é atribuída ao efeito médio dos genes, o que determina o grau de semelhança entre os indivíduos aparentados. A seleção é efetuada, fundamentalmente, baseada nesta semelhança (FALCONER 1987).

DUDLEY e MOLL (1969) relatam que a estimativa de herdabilidade refere-se a uma característica e a uma população e é específica para as condições experimentais nas quais os genótipos foram estudados, sendo inapropriado extrapolar estimativas de herdabilidade de uma população para outra, ou para diferentes condições experimentais.

O modo de herança de uma característica quantitativa pode ser, geralmente, reconhecida de acordo com sua distribuição numa população segregante, tal como populações F2 e RC1F1 e progênies F3 e RC1F2 (GAI e YOU 1994).

O termo característica complexa (quantitativa) se refere a um fenótipo que não exhibe herança Mendeliana clássica, recessivo ou dominante atribuído a um único gene. Em geral, a complexidade surge quando a simples correspondência entre fenótipo e genótipo falha, ou devido ao mesmo genótipo resultar em diferentes fenótipos (efeitos de ambiente, interações com outros genes) ou diferentes genótipos poderem resultar num mesmo fenótipo (LANDER e SCHORK, 1994).

As estimativas de herdabilidade, o ganho esperado com a seleção e as correlações das características selecionadas são extremamente importantes para orientar e planejar programas de melhoramento.

KWON e TORRIE (1964) estimaram herdabilidades de características agronômicas em soja em duas populações derivadas de 64 plantas F2 do cruzamento Norchief x Harosoy e Norchief x Clark. As duas populações foram avaliadas na geração F3 em um local e na geração F4 e F5 em dois anos e dois locais. As estimativas da variância atribuída à interação genótipo x ano foram maiores do que aquelas para interação genótipo x local para muitas características, especialmente produtividade. A

herdabilidade de produção de sementes foi extremamente baixa devido à interação genótipo x ano. As estimativas de herdabilidade de proteína e óleo foram avaliadas em apenas uma das populações (N x H) sendo 57% e 51%, respectivamente.

No trabalho de COBER e VOLDENG (2000) as estimativas de herdabilidade para teor de proteína variaram de 0,40 a 0,77 em linhagens avançadas derivadas de um único ciclo de retrocruzamento e de um único cruzamento.

Como pode ser observada, pelas variações das estimativas de herdabilidade nos vários trabalhos apresentados, esta é uma característica muito influenciada pelo ambiente, no entanto, o progresso na seleção para alto teor de proteína pode ser alcançado.

3.3.2. Melhoria para teor de proteína em soja

O preço da soja é baseado no conteúdo de proteína e óleo na semente. Exportadores de farelo de soja têm reduzido o preço pago para grãos de soja que tenham baixos teores de proteína e óleo (informações obtidas junto a COODETEC).

Os teores de óleo e de proteína são fortemente associados no grão de soja sendo que, em média, o conteúdo de óleo e o conteúdo de proteína somam em torno de 60% do total de peso de matéria seca da semente (JOHNSON *et al.* 1955a e KWON e TORRIE 1964).

Vários programas de melhoria estão empenhados em aumentar o teor de proteína da soja (WILCOX e CAVINS 1995; WEHRMANN *et al.* 1987).

FEHR e WEBER (1968) avaliaram a eficiência de três ciclos de seleção massal pelo tamanho da semente e densidade específica sobre a porcentagem de proteína e óleo em duas populações heterogêneas de soja. Observaram que o progresso máximo foi obtido para alto teor de proteína e baixo teor de óleo quando se selecionou para sementes maiores e de maior densidade específica. O contrário, alto teor de óleo e baixo teor de proteína, foi observado quando se selecionou para sementes pequenas e de baixo

densidade específica. As estimativas de herdabilidade foram de 73 e 77% para óleo e proteína respectivamente.

MILLER e FEHR (1979) avaliaram a eficiência do melhoramento de uma população para conteúdo de proteína em semente de soja por seleção recorrente direta para alto teor de proteína e indireta baseada em baixo teor de óleo. O teor de proteína aumentou de 43,1% para 44,6% no ciclo 0 da população de alto teor de proteína e para 43,9% na população de baixo teor de óleo. A seleção direta causou diminuição significativa no teor de óleo e carboidratos e a seleção indireta diminuiu significativamente o teor de óleo mas não alterou o teor de carboidratos. Estes autores concluíram que a seleção indireta (baseada em baixo teor de óleo) não pode ser considerada um método eficiente para melhoramento do teor de proteína quando o custo e o tempo para análise de proteína não forem fatores limitantes.

Outra estratégia de seleção foi apresentada por ALT *et al.* (2002) que sugerem seleção para alto teor de proteína e semente de tamanho grande em populações obtidas de dois e três cruzamentos. Dois progenitores com alto teor de proteína (AP) e semente de tamanho grande (SG) e um progenitor normal para teor de proteína (N) e tamanho de semente foram usados para obter três tipos de população: APSG X APSG; APSG X N e APSG X (APSG X N). A porcentagem das linhagens F2:3, com tamanho de semente e proteína maior ou igual ao menor progenitor APSG, variou de 4% para o cruzamento APSG X N, 88% para APSG X APSG e 31% para o cruzamento triplo APSG X (APSG X N). Estes autores concluem que as populações dos cruzamentos APSG X APSG e do cruzamento triplo APSG X (APSG X N) seriam mais úteis para desenvolver cultivares com alto teor de proteína e semente de tamanho grande.

ERICKSON, *et al.* (1981) estudaram seleção para proteína em gerações precoce do cruzamento entre *Glycine soja* e *Glycine max*. A herdabilidade do teor de proteína foi medida em F2 e F3 e o ganho por seleção massal e seleção de famílias foram comparados na geração F3 destes cruzamentos. *G. soja* foi usado como progenitor masculino em todos os cruzamentos e a porcentagem de seleção foi de 10%. A herdabilidade para proteína na geração F2 foi de 27% obtida pela regressão pai-filho. A herdabilidade no sentido amplo em famílias F3 cultivadas em dois locais em

uma estação foi de 78%. A média do teor de proteína em cada população selecionada foi maior do que a média da população não selecionada (controle). As médias dos teores de proteína dos métodos de seleção foram: seleção massal em F2 e F3 48%; seleção massal em F3 seguido de uma única semente por descendência, 48,8%; seleção entre famílias F3, 47,6%; seleção entre e dentro de famílias F3, 47,5%. O método de seleção massal foi superior ao método de seleção de famílias em nível de 1%. Estes autores comentam que, apesar de a seleção precoce ter elevado o teor de proteína da população, a segregação continuou a originar genótipos de baixo teor de proteína. Conseqüentemente, passos adicionais de seleção deverão ser necessários.

Tem sido difundido por vários melhoristas que o desenvolvimento de cultivares de soja com alto teor de proteína é limitado por uma correlação negativa entre produtividade e teor de proteína. WILCOX e GUODONG (1997) estudaram a relação entre produtividade e teor de proteína em populações de soja de crescimento determinado e indeterminado. As progênies de crescimento determinado e indeterminado foram oriundas de cruzamentos entre linhagens de alto teor de proteína de crescimento indeterminado com linhagens de crescimento determinado e médio teor de proteína. As análises demonstraram que as progênies de crescimento determinado constituíram melhores fontes de seleções, as quais combinaram alta produtividade e alto teor de proteína, do que as progênies de crescimento indeterminado.

HELMS e ORF (1998) avaliaram o ganho para aumento do teor de proteína por seleção direta e a resposta correlacionada a produtividade e teor de óleo. Seus resultados indicaram que, em média, a seleção para alto teor de proteína resultou em diminuição do teor de óleo e da produtividade.

HARTWIG (1994) demonstrou que progressos podem ser obtidos na seleção de linhagens de soja de alto teor de proteína e alta produtividade quando se utiliza o progenitor de alto teor de proteína com tipo agrônomico melhorado e com produtividade média. Em seu trabalho, uma população F2, com 438 plantas, foi obtida do cruzamento entre a cultivar Forrest (altamente produtiva) com uma linhagem D76-8070 (alto teor de proteína e média produtividade). O teor de proteína das plantas F2 variou de 40,3 a 51,1.

Linhagens com alto teor de proteína e com alta produtividade foram desenvolvidas selecionando 50 plantas com maior teor de proteína em F2. Sete linhagens avançadas em F5 apresentaram teores de proteína similares ao progenitor de alto teor de proteína e produtividade comparável ao progenitor de alta produtividade. Este autor concluiu neste trabalho que a seleção, em gerações precoces para alto teor de proteína, provavelmente, não acarretará nenhum problema no desenvolvimento de cultivares altamente produtivos e com altos teores de proteína.

WILCOX (1998) estudou as mudanças nos teores de proteína e óleo e a variabilidade dessas duas características durante oito ciclos de seleção recorrente em soja. A seleção recorrente aumentou a média do teor de proteína nas sementes de $5,8 \text{ g kg}^{-1}$ e diminuiu a média do teor de óleo de $2,3 \text{ g kg}^{-1}$ por ciclo. Nos últimos ciclos, foi observada uma correlação negativa entre proteína e óleo e diminuição da variabilidade dessas duas características. A maioria dos alelos para alto teor de proteína foi acumulada até o quinto ciclo de seleção recorrente. Uma média de 53% das plantas do sexto até o oitavo ciclo apresentou 480 g kg^{-1} ou mais de proteína. Progênies destas plantas poderiam ser excelentes fontes de fenótipos que combinassem alto teor de proteína com boas características agronômicas.

COBER e VOLDENG (2000) avaliaram melhoramento para alto teor de proteína e produtividade em soja por meio de cruzamento simples e um único ciclo de retrocruzamento. Estes autores concluíram que linhagens avançadas oriundas de um único ciclo de retrocruzamento não foram significativamente diferentes das linhagens avançadas obtidas do cruzamento simples quanto à produtividade e teor de proteína. Ambas as estratégias de melhoramento produziram linhagens com teor de proteína significativamente maior que o progenitor recorrente. Contudo, nenhuma dessas linhagens tiveram produtividade significativamente superior ao progenitor recorrente ou a um cultivar de alto teor de proteína que foi usado como padrão. As populações obtidas exibiram baixa ou nenhuma associação, entre produtividade e teor de proteína. Apesar dos progenitores serem fontes úteis de alelos que não exibiram os efeitos pleiotrópicos de baixa produtividade e alto teor de proteína, concluiu-se que o uso de um ciclo de retrocruzamento para progenitor adaptado não foi eficiente para o

desenvolvimento de linhagens com alto teor de proteína e de alta produtividade.

O farelo de soja contém os açúcares solúveis rafinose e estaquiose os quais são responsáveis pela quantidade limitada do uso do farelo de soja em certas rações. HARTWIG *et al.* (1997) sugerem que um aumento no teor de proteína poderia ser acompanhado de redução no teor de açúcares solúveis, uma vez que com o aumento do teor de proteína e redução do teor de óleo, o teor de proteína total, mais óleo, mais tegumento, é maior do que na soja normal. Sendo essa porção residual reduzida, os açúcares solúveis poderiam fazer parte desta fração que será reduzida e com isso uma fração maior de farelo de soja poderia ser usada em rações.

WILCOX e SHIBLES (2001) estudaram a associação entre teores de proteína, óleo, carboidratos totais, sacarose e enxofre em uma população que continha uma ampla variação para teor de proteína. Neste trabalho, 43 linhagens com concentração de proteína variando de 413 a 468 g kg⁻¹, foram cultivadas em três ambientes. As linhagens apresentaram variações significativas para cada uma das características avaliadas. Os teores de carboidratos totais não tiveram associação com produtividade. O aumento no teor de proteína ocorreu às custas do teor de óleo, carboidratos totais e sacarose com coeficientes de correlação de -0,88, -0,71 e -0,66, respectivamente. No entanto, as proporções S/N foram constantes em todos níveis de proteína. Estes autores concluíram que a diminuição no teor de carboidratos com o aumento do teor proteína deverão contribuir para um aumento do valor nutricional do farelo obtido dessas linhagens melhoradas. Além disto, face à consistente proporção S/N em todas as linhagens com diferentes teores de proteína indica que os teores de aminoácidos sulfurados não foram reduzidos com o aumento de proteína.

A proteína da soja é deficiente em aminoácidos sulfurados, sendo que seu valor nutricional pode ser melhorado, seja pela suplementação artificial com o aminoácido metionina ou por meio do melhoramento genético. Neste sentido, IMSANDE (2001) trabalhou com sementes de soja induzidas por mutação com SEM (Sulfonato de Etil Metano) obtendo linhagens com maior teor de metionina, as quais foram separadas e cruzadas. Sementes de um cruzamento apresentou uma concentração normal de nitrogênio e um

aumento de 18% na concentração de metionina. A proporção S/N da linhagem obtida do cruzamento foi 16,2% maior do que da linhagem parental. As concentrações de metionina e cisteína foram em torno de 20% maior do que aquelas da linhagem parental. Este autor observou que o teor de metionina na proteína da soja foi regulado pela maior ou menor síntese do polipeptídeo β da proteína de reserva β -conglucina, o qual não possui metionina.

3.4. Melhoramento genético de plantas auxiliado por marcadores moleculares

Uma vez que os marcadores moleculares possibilitam análises genéticas mais detalhadas, eles oferecem grandes oportunidades de utilização em programas de melhoramento de plantas. Na introgressão de características de herança simples, os marcadores têm demonstrado vários resultados positivos. Já em relação à seleção assistida por marcadores que envolvem características quantitativas, pouco tem sido feito em termos de aplicação prática (YOUNG 1999).

Atualmente, a utilização de marcadores moleculares em programas de introgressão de genes por retrocruzamento é o exemplo mais concreto de melhoramento genético assistido por marcadores. Germoplasma não-adaptados têm sido utilizados em programas de melhoramento com o objetivo de introduzir características de herança simples, como resistência a patógenos e pragas. O uso de marcadores moleculares ligados a esses genes é de grande importância na seleção de genótipos resistentes, principalmente quando o programa de melhoramento tem como objetivo a introdução de dois ou mais genes de resistência; quando o fenótipo é de determinação complexa; ou quando o processo de avaliação requer a destruição da planta. A progênie de cada ciclo de cruzamento é avaliada com base na presença de uma marca associada à característica de interesse. Este procedimento mascara o efeito de alelos dominantes, elimina a variabilidade devido a efeitos ambientais e pode simplificar padrões de herança para características complexas (RAFALSKI *et al.* 1993). Além de monitorar a presença do gene de interesse, a genotipagem dos indivíduos

com marcadores moleculares, permite a seleção dos indivíduos mais semelhantes ao progenitor recorrente e com melhor conversão na região próxima ao gene introduzido (VISSCHER *et al.* 1996; HOSPITAL *et al.* 1992). Deste modo, o número de ciclos de retrocruzamentos necessários para a recuperação do fenótipo recorrente é reduzido de forma acentuada, acelerando o desenvolvimento de variedades melhoradas (OPENSHAW *et al.* 1994).

Muitos dos problemas encontrados em um programa de melhoramento baseado na estimativa fenotípica de uma característica agrônômica tais como efeitos ambientais ou herança quantitativa podem ser eliminados por diagnósticos baseados na análise do DNA (RAFALSKI *et al.* 1993).

O reconhecimento de que marcadores baseados em DNA podem ser extremamente úteis no melhoramento de plantas foi obtido com o desenvolvimento de mapeamento genético de plantas baseado em RFLP (KEIM *et al.* 1989; BERNATZKY *et al.* 1986). Como já mencionado anteriormente, marcadores moleculares de DNA que flanqueiam um gene que determina uma característica podem ser usados para seguir a característica em cruzamentos genéticos. Além disto, marcadores moleculares que são uniformemente distribuídos por todo o genoma podem ser usados para estimar a contribuição genética de cada parental para cada membro de uma população segregante (VANTOAI *et al.* 1997; LORENZEN *et al.* 1996). Indivíduos cuja composição genômica mais se aproxima do progenitor recorrente (retrocruzamento) são selecionados e usados para o próximo cruzamento (OPENSHAW *et al.* 1994). Isto permite acelerar a introgressão de características de fontes de germoplasma exóticas.

Os primeiros trabalhos de mapeamento com marcadores moleculares utilizaram caracteres de herança simples, pois as variações fenotípicas determinadas pela segregação de um ou poucos genes são fáceis de serem caracterizadas. Linhagens isogênicas ou isolinhas têm sido bastante usadas nesses estudos. Estas linhagens são desenvolvidas a partir de ciclos sucessivos de retrocruzamentos com o progenitor recorrente, tendo como objetivo introduzir um gene de interesse em uma linhagem comercial, originando uma linhagem geneticamente semelhante à original com exceção

da região genômica onde o gene foi introduzido. A análise de isolinhas com marcadores moleculares permitiu a identificação de marcadores ligados a genes de interesse em várias espécies de importância econômica, como resistência ao vírus TMV em tomate (YOUNG *et al.* 1988), resistência a *Pseudomonas* em tomate (MARTIN *et al.* 1991), resistência a *Pyricularia grisea* em arroz (YU *et al.* 1991) e resistência a *Phytophthora megasperma* em soja (DIERS *et al.* 1992).

ARNHEIM *et al.* (1985) analisaram misturas de DNA de indivíduos homozigóticos para locos de interesse, com o objetivo de identificar marcadores associados a esses locos na espécie humana. Utilizando este mesmo princípio em plantas, MICHELMORE *et al.* (1991) desenvolveram a técnica de "bulks" segregantes - BSA (*Bulked Segregant Analysis*), que tem revolucionado a identificação de regiões genômicas associadas a caracteres de herança simples. Este trabalho identificou marcadores RAPD ligados a genes de resistência ao míldio pulverulento em alface, sendo o marco inicial para a ampla difusão da técnica de BSA. Esta metodologia baseia-se na construção de dois "bulks" de DNA contrastantes para uma característica fenotípica entre os indivíduos de uma população segregante. Cada "bulk" contém quantidades iguais de DNA de indivíduos com fenótipos extremos, que compartilham uma mesma região genômica que contém o gene de interesse. O DNA dos dois "bulks" contrastantes é testado com marcadores e os polimorfismos encontrados têm uma grande chance de estarem ligados à característica que contrasta os "bulks", uma vez que as demais características segregam aleatoriamente. A ligação dos marcadores obtidos com o fenótipo desejado é testada pela análise de co-segregação na população segregante. Uma das vantagens do BSA é que a identificação de marcadores ligados a caracteres de interesse não requer a construção de um mapa genético saturado.

A técnica do BSA tem sido utilizada com sucesso na identificação de marcadores ligados a características de herança simples, como no caso de genes de resistência a doenças. Inúmeros exemplos são encontrados na literatura: identificação de marcadores RAPD e RFLP associados à resistência ao míldio (*Bremia lactucae*) em alface (MICHELMORE *et al.* 1991; PARAN *et al.* 1991; PARAN e MICHELMORE 1993); identificação de

marcadores RAPD ligados ao gene *Pg3*, que confere resistência a *Puccinia graminis* em aveia (PENNER *et al.* 1993); e marcadores ligados à resistência a *Uromyces phaseoli* em feijão (MIKLAS *et al.* 1993; HALEY *et al.* 1993). ALZATE-MARIN (1996) identificou marcadores RAPD ligados em fase de acoplamento e de repulsão a genes de resistência a *Colletotrichum lindemuthianum*, agente causal da antracnose do feijoeiro. De forma similar, usando a análise de “bulks” segregantes, foi identificado um marcador RAPD ligado ao gene que confere resistência ao cancro da haste da soja (CARVALHO *et al.* 2002). Regiões genômicas que controlam outras características, além da resistência a patógenos, como amadurecimento do fruto e abscisão do pedicelo em tomate, foram também mapeadas utilizando essa estratégia (GIOVANNONI *et al.* 1991).

Os tipos de marcadores moleculares mais usados no melhoramento são o RAPD (*Randomly Amplified Polymorphic DNA*) e microssatélites (*Simple Sequence Repeats* – SSR). Ambos marcadores, baseados na técnica de reação em cadeia da polimerase (*Polimerase Chain Reactions* - PCR), constituem ferramentas muito úteis e práticas, por serem tecnicamente simples e de fácil montagem, requererem poucas quantidades de DNA e permitem que dezenas de plantas possam ser analisadas de uma só vez (WAUGH e POWELL 1992).

Os marcadores RAPD (Random Amplified of Polymorphic DNA) (WILLIAMS, *et al.*, 1990) ou AP-PCR (Arbitrarily Primed-PCR) (WELSH e McCLELLAND 1990) têm sido amplamente utilizados devido a sua simplicidade e fácil uso em laboratórios equipados modestamente (RAFALSKI e TINGEY 1993).

O RAPD é uma variação da técnica de PCR que utiliza um único *primer* de dez nucleotídeos e com seqüência arbitrária. Portanto, para que um fragmento de DNA seja amplificado, duas regiões complementares ao *primer* devem estar separadas por até 2.000 pb e em orientações opostas na fita de DNA. Com isso, são amplificados fragmentos de DNA distribuídos ao acaso no genoma sem conhecimento prévio da seqüência de DNA. A detecção dos produtos de amplificação é feita em gel de agarose contendo brometo de etídio e visualizados sob luz ultravioleta. As bases moleculares do polimorfismo de RAPD são mutações de ponto ou deleções no sítio de

pareamento do *primer*, ou inserções entre os sítios de pareamento, produzindo uma distância que impossibilita a sua amplificação. As desvantagens da técnica são a reprodutibilidade dos padrões de amplificação e o baixo conteúdo de informação por loco, uma vez que são marcadores dominantes.

Por suas características, marcadores moleculares baseados na variação do comprimento de seqüências simples repetidas (SSR) ou seqüências de DNA microssatélites têm sido um dos mais usados no melhoramento de plantas. Microssatélites consistem de unidades núcleo de 2-5 nucleotídeos tais como (CA), (ATT) ou (ATGT) que são repetidas em tandem no genoma (LITT e LUTY 1989). As regiões que flanqueiam um microssatélite são geralmente conservadas entre genótipos da mesma espécie e *primers* são construídos para estas regiões e utilizados para amplificar fragmentos de DNA contendo o microssatélite. Polimorfismo de comprimento é criado quando produtos de PCR de diferentes alelos variam no comprimento, como resultado da variação do número de unidades repetidas no microssatélite, podendo, então, serem analisados por eletroforese em gel de acrilamida ou gel de agarose. O alto nível de informação e a co-dominância de marcadores microssatélites, sua grande ocorrência em genomas eucarióticos e sua fácil amplificação via tecnologia padrão de PCR, tornou os microssatélites marcadores preferidos para várias espécies (CREGAN *et al.* 1994; RUSSEL *et al.* 1997). Neste sentido, marcadores microssatélites têm sido amplamente utilizados na caracterização e avaliação da diversidade genética de genótipos de soja (ROGWEN *et al.* 1995; NETO *et al.* 2001; NARVEL *et al.* 2000; TANYA *et al.* 2001).

REFERÊNCIAS BIBLIOGRÁFICAS

- ALZATE-MARIN, A.L. **Resistência à antracnose do feijoeiro (*Phaseolus vulgaris* L.): diversidade genética de raças de *Colletotrichum lindemuthianum*, herança da resistência e identificação de marcadores moleculares.** Viçosa, MG:UFV, 1996. 65p. Dissertação (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa. 1996.
- ARNHEIM, N., STARNGE, C. ERLICH, H. Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: studies of HLA class II loci. **Proc. Natl. Acad. Sci. USA**, v.82, p.6970-6974, 1985.
- ASSOCIATION OF OFFICIAL ANALYTICAL CHEMISTS – AOAC. **Official methods of analysis.** Washington, D.C.: 1975. 1094p.
- ABDELNOOR, R.V., BARROS, E.G., MOREIRA, M.A. Determination of genetic diversity within brazilian soybean germplasm using random amplified polymorphic DNA techniques and comparative analysis with pedigree data. **Brazil. J. Genetics**, v.18, p.265-273, 1995.
- BERNATZKY, R., TANKSLEY, S.D. Toward a saturated linkage map of tomato based on isozymes and random cDNA sequences. **Genetics**, v.112, p.887-898, 1986.
- BURTON, J.W. Soybean (*Glycine max* L. Merr.). **Field crops research**, v. 53, p. 171-186, 1997.
- ALT, B.J., FEHR, W.R., WELKE, G.A. Selection for large seed and high protein in two and three parent soybean populations. **Crop Sci.**, v.42, p.1876-1881, 2002.
- CARVALHO, G.A., SEDIYAMA, T., MARIN, A.L.A., BARROS, E.G., MOREIRA, M.A. Identificação de marcadores RAPD ligados a um gene de resistência ao cancro da haste da soja. **Fitopatologia Brasileira**, v.27, n.3, p.474-478, 2002.
- COBER, E.R., VOLDENG, H.D., Developing high protein, high-yield soybean populations and lines. **Crop Sci.** v.40, n.1, p.39-42, 2000.

- DIERS, B.W., MANSUR, L., IMSANDE, J. SHOEMAKER, R.C. Mapping *Phytophthora* resistance loci in soybean with restriction fragment length polymorphism markers. **Crop Sci.**, v.32, p.377-383, 1992.
- CREGAN, P.B., BHAGWAT, A.A., AKKAYA, M.S., RONGWEN, J. Microsatellite fingerprinting and mapping of soybean. **Meth. Mol. Cell. Biol.** v.5, p.49-61, 1994.
- DOYLE, J.J, DOYLE, J.L. Isolation of plant DNA from fresh tissue. **Focus**, v.12, p.13-15, 1990.
- DUDLEY, J.W., MOLL, R.H. Interpretation and use of estimates of heritability and genetics variances in plant breeding. **Crop Sci.**, v.9, n.3, p.257-262, 1969.
- ERICKSON, L.R., VOLDENG, H.D., BEVERSDORF, W.D. Early generation selection for protein in *Glycine max* X *glycine soja* crosses. **Can. J. Plant Sci.** vol.61, p.901-908, 1981.
- FALCONER, D.S. **Introdução à genética quantitativa**. Viçosa, MG: UFV,1987. 279p.
- FEHR, W.R., WEBER, C.R. Mass selection by seed size and specific gravity in soybean populations. **Crop Sci.**, v.8, p.551-554, 1968.
- GAI, J., YOU, M. Genetics of quantitative traits of soybeans. In: **Genetic improvement. Proceedings: World Soybean Reseach Conference V.** Kasetsart University Press and WSRCV, p.74-83, 1994.
- GIOVANNONI, J.J., WING, R.A., GANAL, M.W., TANKSLEY, S.D. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. **Nucl. Acids Res.**, v.19, p.6553-6558, 1991.
- HALEY, S.D., MIKLAS, P.N., STAVELY, J.R., BYRUM, J., KELLY, J.D. Identification of RAPD marker linked to a major rust resistance gene block in common bean. **Theor. Appl. Genet.**, v.86, p.505-512, 1993.
- HARTWIG, E.E. Breeding soybean for high seed yield and seed protein. In: **Genetic Improvement. Proceedings: World Soybean Reseach Conference V.** Kasetsart University Press and WSRCV, p.74-83, 1994.

- HARTWIG, E.E., KUO, T.M., KENTY, M.M. Seed protein and its relationship to soluble sugars in soybean. **Crop Sci.**, v.37, p.770-773, 1997.
- HELMS, T.C., ORF, J.H. Protein, oil and yield of soybean lines selected for increased protein. **Crop Sci.**, v.38, p.707-711, 1998.
- HILL, J.E., BREIDENBACH, R.W. Proteins of soybean seeds. II. Accumulation of the major protein components during seed development and maturation. **Plant physiol.**, v.53, p.747-751, 1974.
- HIROMOTO, D.M., VELLO, N.A. The genetic base os brazilian soybean (*glycine Max* (L.) Merrill) cultivars. **Rev. Brasil. Genet.** v.IX, n.2 p.295-306, 1986.
- HOSPITAL, F., CHEVALET, C., MULSANT, P. Using markers in gene introgression breeding programs. **Genetics**, v.132, p.1199-1210, 1992.
- HUGHES, S., MURPHY, P. Varietal influence on the quantity of glycinin in soybeans. **J. Agric. Food Chem.**, v.31, p.376-379, 1983.
- IMSANDE, J. Selection of soybean mutants with increased concentrations of seed methionine and cysteine. **Crop Sci.** vol. 41, p.510-515, 2001.
- INSTITUTO ADOLFO LUTZ. **Normas analíticas do Instituto Adolfo Lutz, Métodos químicos e físicos para análise de alimentos.** 3° ed., São Paulo, v.1, p.533, 1985.
- JOHNSON, H.W., ROBINSON, H.F., COMSTOCK, R.E. Genotypic and phenotypic correlations in soybean and their implications in selection. **Agron. J.**, v.47, p.477-483, 1955b.
- KEIM, P., SHOEMAKER, R., PALMER, R. Restriction fragment length polymorphism diversity in soybean. **Theor. Appl. Genet.**, v.77, p.786-792, 1989.
- KITAMURA, K., TOYOKAWA, Y., HARADA, K. Polimorphism of glycinin in soybean seeds. **Phytochemistry**, v.19, p.1841-1843, 1980.
- KITAMURA, K., KAIZUMA, N. Mutants strains with low level of subunits of 7S globulin in soybean (*glycine max* Merr.) seed. **Japan. J. Breed.**, v.31, p.353-359, 1981.
- KITAMURA, K., DAVIES, C.S. NIELSEN, N.C. Inheritance of alleles for Cgy₁ and Gy₄ storage protein genes in soybean. **Theor. Appl. Genet.**, v.68, p.253-257, 1984.

- KITAMURA, K. Genetic improvement of nutritional and food processing quality in soybean. **JARQ**, v.29, p.1-8, 1995.
- KOSHIYAMA, I. Storage proteins of soybean. In: **Soybean: Improvement, production and uses**. Rd., Madison, WI, USA, 2nd ed., Agronomy monograph, n.16, p.427-450, 1987.
- KWON, S.H., TORRIE, J.H. Heritability of and interrelationships among traits of two soybean populations. **Crop Sci.** v. 12, n.2, p.196-198, 1964.
- LANDER, E.S., SCHORK, N.J. Genetic dissections of complex traits. **Science**, v.265, p.2037-2048, 1994.
- LITT, M. LUTY, J.A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. **Am. J. Hum. Genet.** v.44, p.397-401, 1989.
- LORENZEN, L.L., BOUTIN. S., YOUNG, N., SPECHT, E., SHOEMAKER, R.C. Soybean pedigree analysis using map-based molecular markers: I. Tracking RFLP markers in cultivars. **Crop. Sci.**, v.35, p.1326-1336, 1995.
- LORENZEN, L.L., LIN, S.F., SHOEMAKER, R.C. Soybean pedigree analysis map-based molecular markers: recombination during cultivar development. **Theor. Appl. Genet.**, v.93, p.1251-1260, 1996.
- MARTIN, G.B., WILLIAMS, J.G.K., TANKSLEY, S.D. Rapid identification of markers linked to a *Pseudomonas* resistance genes in tomato using random primer and near-isogenic lines. **Proc. Natl. Acad. Sci. USA**, v.88, p.2336-2340, 1991.
- MIKLAS, P.N., STAVELY, J.R., KELLY, J.D. Identification and potential use of a molecular marker for rust resistance in common bean. **Theor. Appl. Genet.**, v.85, p.745-749, 1993.
- MILLER, J.E., FEHR, W.R. Direct and indirect recurrent selection for protein in soybeans. **Crop Sci.**, v.19, p.101-106, 1979
- MICHELMORE, R.W., PARAN, I., KESSELI, R.V. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. **Proc. Natl. Acad. Sci. USA**, v.88, p.9828-9832, 1991.

- MOREIRA, M.A., HERMODSON, M.A., LARKINS, B.A., NIELSEN, N.C. Partial characterization of the acidic and basic polypeptides of glycinin. **J. Biol. Chem.** v.254, p.9921-9926, 1979.
- MOREIRA, M.A., HERMODSON, M.A., LARKINS, B.A., NIELSEN, N.C. Comparison of the structure of the acidic polypeptides of glycinin. **Arch. Of Bioch. and Byoph.**, v.210, p.636-642, 1981.
- MORI, T., UTSUMI, S., INABA, H., KITAMURA, K., HARADA, K. Differences in subunit composition of glycinin among soybean cultivars. **J. Agric. Food Chem.**, v.29, p.20-23, 1981.
- MORI, T., UTSUMI, S., INABA, H., **Agric. Biol. Chem.** V.43, p.2317, 1979.
- MURPHY, P.A. RESURRECION, P. Varietal and environmental differences in soybean glycinin and β -glycinin content. **J. Agric. Food Chem.** V.32, p.911-915, 1984.
- NARVEL, J.M., FEHR, W.R., CHU, W.C., GRANT, D. SHOEMAKER, R.C. Simple sequence repeat diversity among soybean plant introductions and elite genotypes. **Crop Sci.**, v.40, n.5, p.1452-1458, 2000.
- NETO, F.A. **Marcadores microssatélites na identificação de cultivares de soja.** Viçosa, MG:UFV, 2001. 46p. Dissertação (Mestrado em Fitotecnia) - Universidade Federal de Viçosa, Viçosa. 2001.
- OGAWA, T., TAYAMA, KITAMURA, K., KAIZUMA, N. Genetic improvement of seed storage proteins using three variant alleles of 7S globulin subunits in soybean (*glycine max* L.). **Japan J. Breed.**, v.39, p.137-147, 1989.
- OPENSHAW, S.J., JARBOE, S.G., BEAVIS, W.D. Marker-assisted selection in backcross breeding. In: **ASHS/CSSA Joint Plant Breeding Symposium, 2, Corvallis Proceedings...** Corvallis: Oregon State University, 1994.
- PALMER, R.G., HYMOWITZ, T., NELSON, R.L. Germoplasm diversity within soybean. In: **Soybean genetics, molecular biology and biotechnology.** Ohio, USA:CAB INTERNACIONAL, p.1-35, 1996.
- PARAN, I., KESSELI, R., MICHELMORE, R.W. Identification of restriction fragment length polymorphism and random amplified polymorphic DNA markers linked to downy mildew resistance genes in lettuce, using near-isogenic lines. **Genome**, v.34, p.1021-1027. 1991.

- PARAN, I., MICHELMORE, R.W. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. **Theor. Appl. Genet.**, v.85, p.985-993, 1993.
- PENNER, G.A., CHONG, J., LEVESQUE-LEMAY, M., MOLNAR, S.J. FEDAK, G. Identification of a RAPD marker linked to the oat stem rust gene *Pg3*. **Theor. Appl. Genet.**, v.85, p.702-705, 1993.
- PAULS, K.P. Plant biotechnology for crop improvement. **Biotechnology Advances**, vol.13, n.4, p.673-693, 1995.
- RAFALSKI, J.A., SCOTT, V.T. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. **Trends in Genetics**, v.9, n.8, p.275-280, 1993.
- ROGWEN, J., AKKAYA, M.S., BHAGWAT, A.A., LAVI, U., CREGAN, P.B. The use of microsatellite DNA markers for soybean genotypes identification. **Theor. Appl. Genet.**, v.90, n.1, p.43-48, 1995.
- RUSSELL, J., FULLER, J., YOUNG, G., THOMAS, B., TARAMINO, G., MACAULAY, M., WAUGH, R., POWELL, W. Discriminating between barley genotypes using microsatellites markers. **Genome**, v.40, p.442-450, 1997.
- STASWISCK, P.E., NIELSEN, N.C. Characterization of a soybean cultivar lacking certain glycinin subunits. **Arch. Bioch. Bioph.**, v.223, n.1, p.1-8, 1983.
- SMITH, H., GRIERSON, D. Seed maturation and deposition of storage proteins. In: The molecular biology of plant development, SMITH, H., E GRIERSON, D (ed) **Botanical monograph**, University of California Press, Berkeley, v.18, p.306-336, 1982.
- SMITH, P.K.; KROHN, R.I.; HERMANSON, G.T.; MALLIA, A. K.; GARTNER, F. H.; PROVENZANO, M.D.; FUJIMOTO, E.K.; GOEKE, M.M.; OLSON, B.J. KLENK, D. C. Measurement of protein using bicinchoninic acid. **Anal. Biochem.**, v.15, p.76-85, 1985.
- TAKAHASHI, K., MIZUNO, Y., YUMOTO, S., KITAMURA, K., NAKAMURA, S. Inheritance of the α -subunit deficiency of β -conglycinin in soybean (*glycine max* L. Merrill) line induced by γ -ray irradiation. **Breeding Science**, v.46, n.3, p.251-255, 1996.

- TANYA, P., SRINIVES, P., TOOJINDA, T., VANAVICHIT, A., HA, B.K., BAE, J.S., MOON, J.K., LEE, S.H. Evaluation of genetic diversity among soybean genotypes using SSR and SNP. **Korean J. Crop Sci.**, v.46, n.4, p.334-340, 2001.
- VERMA, D.P.S., SHOEMAKER, R.C. Prefácio. In: **Soybean genetics, molecular biology and biotechnology**. Ohio, USA: CAB INTERNACIONAL, 1996. p.ix.
- VANTOAI, T.T., PENG, J., ST. MARTIN, S.K. Using AFLP to determine the genomic contribution of parents to populations. **Crop Sci.**, v.37, p.1370-1373, 1997.
- TAUTZ, D., RENZ, M. Simple sequences are ubiquitous repetitive components of eukaryote genomes. **Nucl. Acids Res.**,v.12, p.4127-4137, 1984.
- UTUMI, M.M., BARROS, E.G., OLIVEIRA, M.G.A., SEDIYAMA, C.S., MOREIRA, M.A. Efeito da eliminação genética de lipoxigenase e de polipeptídeos de reserva na qualidade da proteína de soja. **Revista Brasileira de Fisiologia Vegetal**. v.10, n.3, p.203-212, 1998.
- VISSCHER, P.M., HALEY, C.S., THOMPSON, R. Marker assisted introgression in backcross breeding programs. **Genetics**, v.144, p.1923-1932, 1996.
- WELSH, J., McCLELLAND, M. Fingerprinting genomes using PCR with arbitrary primers. **Nucl. Acids Res.**, v.18, p.7213-7218, 1990.
- WEHRMANN, V.K., FEHR, W.R., CIANZIO, S.R., CAVINS, J.F. Transfer of high seed protein to high-yielding soybean cultivars. **Crop Sci.** v. 27, p.927-931, 1987.
- WILCOX, J.R., CAVINS, J.F., Backcrossing high protein to a soybean cultivar. **Crop Sci.** v. 35, p.1036-1041, 1995.
- WILCOX, J.R., GUODONG, Z. Relationship between seed yield and protein in determinate and indeterminate soybean population. **Crop Sci.** v. 37, p.361-364, 1997.
- WILCOX, J.R. Increasing seed protein in soybean with eight cycles of recurrent selection. **Crop Sci.** v. 38, p.1536-1540, 1998.
- WILCOX, J.R., SHIBLES, R.M. Interrelationships among seeds quality attributes in soybean. **Crop Sci.** v. 41, p.11-14, 2001.

- WILLIAMS, J.G.K., KUBELICK, A.R., LIVAK, K.J., RAFALSKI, J.A. TINGEY, S.V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. **Nucl. Acids Res.**, v.18, p.6531-6535, 1990.
- YOUNG, N.D. Cautiously optimistic vision for marker assisted breeding. **Molecular Breeding**, v.5, p.505-510, 1999.
- YOUNG, N.D., TANSLEY, S.D. Restriction fragment length polymorphism maps and concept of graphical genotypes. **Theor. Appl. Genet.**, v.77, p.95-101, 1988.
- YU, Z.H., MacKILL, D.J., BONMAN, J.M., TANKSLEY, S.D. Tagging genes for blast resistance in rice via linkage to RFLP markers. **Theor. Appl. Genet.**, v.81, p.471-476, 1991.

CAPÍTULO 1

DIVERGÊNCIA GENÉTICA ENTRE PROGENITORES DE SOJA PARA USO EM PROGRAMAS DE RETROCRUZAMENTOS

RESUMO

Em geral, os progenitores de cruzamentos são selecionados com base em seus desempenhos e na divergência genética entre eles. Em programas de retrocruzamentos, a seleção de indivíduos geneticamente mais próximos do progenitor recorrente reduz o número de retrocruzamentos necessários para a recuperação do genoma do progenitor recorrente. Nesse sentido, análises realizadas com marcadores microssatélites são extremamente úteis, pois estes estão distribuídos ao longo do genoma, resultando em um melhor acompanhamento da introgressão de genes durante o processo de retrocruzamentos. Este trabalho teve como objetivo selecionar indivíduos geneticamente mais próximos entre si, com base em suas similaridades genéticas estimadas pelo coeficiente de coincidência simples e agrupados pelos métodos UPGMA, Vizinho Mais Próximo e de Tocher. Três variedades comerciais e seis linhagens de soja com teor normal de proteína da Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda (COODETEC) e dezoito linhagens doadoras de alelos para alto teor de proteína do Programa de Melhoramento da Qualidade da Soja do BIOAGRO/UFV foram avaliadas quanto às suas diversidades genéticas utilizando 57 pares de *primers* de microssatélites. As distâncias genéticas variaram de 0,08 a 0,74. Os resultados obtidos possibilitaram separar os progenitores com base em suas distâncias genéticas, permitindo que genótipos mais próximos entre si fossem utilizados nos programas de retrocruzamentos objetivando a introdução de alelos de alto teor de proteína em variedades elite de soja. O método UPGMA permitiu a formação de cinco grupos de progenitores, enquanto que o do Vizinho Mais Próximo seis grupos e o de Tocher nove grupos. Comparando os três métodos de agrupamento, pode-se concluir que o método de Tocher foi o que proporcionou a maior diferenciação entre os progenitores, produzindo o maior número de grupos, no entanto, os três métodos foram bastante coincidentes entre si para a maioria dos grupos.

Palavras chave: *fingerprinting*, microssatélites, distâncias genéticas, marcadores moleculares

INTRODUÇÃO

A divergência genética, quantificada pela diferença nas frequências alélicas das populações, entre um grupo de progenitores tem sido avaliada com objetivo de identificar as combinações híbridas de maior efeito heterótico e maior heterozigose, de tal forma que, nas gerações segregantes, se tenha maior possibilidade de seleção de genótipos superiores. Em geral, os progenitores para cruzamentos são selecionados com base em seus desempenhos e na divergência genética entre eles (CRUZ 1997).

Estas avaliações têm sido realizadas por meio de técnicas biométricas baseadas na quantificação da heterose, como nas análises dialélicas, e por meio de processos preditivos. No entanto, para utilização das análises dialélicas, é necessária a avaliação de cada progenitor em todas as combinações possíveis, o que pode inviabilizar o estudo no caso da existência de muitos progenitores. Os métodos preditivos, por outro lado, por dispensarem a obtenção prévia das combinações híbridas, têm merecido considerável ênfase. Estes métodos tomam por base as diferenças morfológicas, fisiológicas e moleculares, entre outras, apresentadas pelos progenitores na avaliação da divergência, que é geralmente quantificada por uma medida de dissimilaridade. A escolha do método mais adequado tem sido determinada em função da precisão desejada, da facilidade da análise e pela forma como os dados foram obtidos (CRUZ 1997).

Assim que mapas genéticos moleculares começaram a surgir (SHOEMAKER *et al.* 1992; SHOEMAKER e SPECHT 1995; MORGANTE 1994; AKKAYA *et al.* 1995; BERNATZJY *et al.* 1986), os marcadores moleculares baseados em mapas tornaram úteis no melhoramento genético como uma poderosa ferramenta. O termo melhoramento molecular é usado para descrever programas de melhoramento que são auxiliados por diagnósticos baseados em DNA. Uma das classes de marcadores que tem sido mais utilizada no melhoramento é a classe dos microssatélites ou SSR (*Simple Sequence Repeats*) que são seqüências de 1 a 6 pb repetidas ao longo do genoma. SSRs são amplificados por PCR com *primers* que são complementares às regiões conservadas que flanqueiam uma região SSR.

O polimorfismo dos fragmentos (alelos) resulta da variação no comprimento das repetições SSR que são separados em gel de agarose ou poliacrilamida (NARVEL 2000). Devido à ampla distribuição dos marcadores SSR ao longo dos cromossomos, estes permitem a avaliação mais detalhada de todo o genoma, resultando em um melhor acompanhamento da introgressão de genes durante o processo de retrocruzamento. Além disto, estes marcadores fornecem subsídios na seleção de populações básicas a serem utilizadas em programas de melhoramento, gerando informações de agrupamento de genótipos e planejamento de cruzamentos (BROWN-GUEDIRA *et al.* 2000). Dependendo do número de indivíduos avaliados as estimativas de distâncias genéticas tornam-se muito difíceis de serem manuseadas e avaliar o relacionamento entre indivíduos dificulta o trabalho de seleção. Para contornar esse problema, faz-se uso de procedimentos estatísticos como a análise de agrupamentos, construção de dendrogramas, estimativa de componentes principais ou ainda o uso de escala multidimensional (SANTOS 1994). Este trabalho teve como objetivo selecionar progenitores de soja geneticamente mais próximos entre si, com base em suas dissimilaridades genéticas determinadas pela análise com microssatélites, estimadas pelo coeficiente de coincidência simples e agrupados pelos métodos UPGMA, Vizinho Mais Próximo e de Tocher, visando a introgressão de alelos para alto teor de proteína em variedades elite de soja por retrocruzamentos.

MATERIAL E MÉTODOS

Material genético e cruzamentos

Três variedades (CD 204, CD 210, CD 211) e seis linhagens de soja (CD 983321, CD 985015, CD 983343, OC 953006, OC 953194, e OC 953312) provenientes da Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda - COODETEC constantes da Tabela 1 foram utilizadas como progenitores recorrentes. Dezoito linhagens, com alto teor de proteína do programa de melhoramento da soja do BIOAGRO/UFV, foram utilizadas como fontes doadoras da característica alto teor de proteína com o objetivo de se incorporar alelos para alto teor de proteína nas variedades da COODETEC.

Tabela 1 - Progenitores usados no programa de retrocruzamentos visando a incorporação de alelos para alto teor de proteína em variedades de soja da COODETEC.

Progenitor	Genótipo	Categoria
1	OC 128PT258	Doador
2	OC 14PT260	Doador
3	CD 203PT30-1	Doador
4	CD 203PT30-3	Doador
5	CD 203PT30-2	Doador
6	CD 202PT31-3	Doador
7	CD 202PT31-1	Doador
8	CD 202PT31-2	Doador
9	CD 202PT111-1	Doador
10	OC 671PT100-1	Doador
11	OC 671PT46-1	Doador
12	OC 671PT46-2	Doador
13	OC 13PT158-2	Doador
14	OC 13PT158-3	Doador
15	CD 983321	Recorrente
16	CD 985015	Recorrente
17	CD 201PT82-3	Doador
18	OC 953194	Recorrente
19	CD 211	Recorrente
20	CD 210	Recorrente
21	OC 953006	Recorrente
22	OC 953312	Recorrente
23	CD 983343	Recorrente
24	CD 204	Recorrente
25	B1PTA25	Doador
26	B1PTA97	Doador
27	B1PTA217	Doador

Extração e purificação de DNA

Amostras de DNA das sementes foram extraídas pelo método descrito segundo MCDONALD *et al.* (1994) com algumas modificações. Em tubo eppendorff de 1,5 mL contendo cerca de 50 mg de sementes eram adicionados 1000 µL de tampão de extração contendo Tris-HCl 0,2 M, NaCl 0,28 M, EDTA 25 mM e SDS 10%. As amostras eram extraídas em Politron e, em seguida, centrifugadas por 10 min por 14.000 RPM em microcentrífuga eppendorff 1405. Os sobrenadantes eram transferidos para novos tubos, sendo, em seguida, acrescentados 10 µL de proteinase K 10mg/mL e 10 µL de CaCl₂ 1 mM, e colocado em banho-maria a 55 °C por 30 min. Eram, então, adicionados às amostras 900 µL de isopropanol e deixadas em repouso por 2 min. Após este tempo as amostras eram centrifugadas por 10 min a 14.000 RPM. Os sobrenadantes eram descartados e os precipitados secos por 15 min à temperatura ambiente sendo, a seguir, ressuspensos em solução de Tris-EDTA (TE) 10 mM e 1 mM contendo RNase 60 µg/mL e colocados em banho-maria por 90 min. As amostras eram precipitadas novamente pela adição de 900 µL de isopropanol e deixadas em precipitação por 2 min. Logo após as amostras eram novamente centrifugadas por 10 min a 14.000 RPM e os sobrenadantes eram descartados. Os precipitados eram ressuspensos ao final em TE 10 mM e 1 mM. A concentração do DNA era estimada espectrofotometricamente por leitura da absorbância em 260 nm, sendo cada unidade de absorbância correspondente a 50 µg/mL de DNA fita dupla (SAMBROOK *et al.* 1989).

Reações de amplificação

As reações de amplificação de microssatélites eram feitas em um volume total de 15 µL, contendo 12,5 µM de Tris-HCl (pH 8,3), 62,5 µM de KCl, 2,5 mM de MgCl₂, 125 µM de cada um dos deoxinucleotídios (dATP, dTTP, dGTP e dCTP), 0,6 µM de cada *primer*, uma unidade da enzima Taq-polimerase e 18 ng de DNA. As amplificações eram realizadas em Termociclador Perkin Elmer 9600 programado para uma etapa inicial de 7 min. a 72 °C, seguida de 30 ciclos de 1 min. a 94 °C, 1 min. a 50 °C e 2 min.

a 72 °C. Por fim, uma etapa de 7 min. a 72 °C. Os fragmentos amplificados eram separados por eletroforese em gel não desnaturante de poliacrilamida 10% por 4 horas a 100 V, e corados com solução de brometo de etídio (0,2 µg/mL). O gel possuía 16,5x12,5 cm e espessura de 0,5 mm. Terminada a eletroforese, os géis eram fotodigitalizados pelo sistema de fotodocumentação Eagle Eye II (Stratagene).

Seleção de pares de primers de microsatélites

De um total de cerca de 600 pares de *primers* de microsatélites disponíveis para soja, foram testados 86 e selecionados 57 com base no mapa integrado da soja (CREGAN *et al.* 1999), os quais se mostraram polimórficos para obter os dados moleculares dos genótipos, sendo eles listados a seguir: (1)Sat009, (2)Satt005, (3)Satt066, (4)Satt070, (5)Satt079, (6)Satt082, (7)Satt115, (8)Satt131, (9)Satt135, (10)Satt136, (11)Satt146, (12)Satt157, (13)Satt163, (14)Satt173, (15)Satt179, (16)Satt181, (17)Satt183, (18)Satt186, (19)Satt187, (20)Satt197, (21)Satt200, (22)Satt212, (23)Satt215, (24)Satt235, (25)Satt238, (26)Satt242, (27)Satt250, (28)Satt260, (29)Satt262, (30)Satt263, (31)Satt269, (32)Satt270, (33)Satt275, (34)Satt290, (35)Satt304, (36)Satt302, (37)Satt309, (38)Satt311, (39)Satt314, (40)Satt320, (41)Satt322, (42)Satt323, (43)Satt353, (44)Satt358, (45)Satt354, (46)Satt362, (47)Satt373, (48)Satt380, (49)Satt381, (50)Satt385, (51)Satt415, (52)Satt512, (53)Satt527, (54)Satt536, (55)Satt545, (56)Satt551 e (57)Satt574.

Estimadores das distâncias genéticas

As distâncias genéticas entre os progenitores foram estimadas com base nos dados das análises feitas com os microsatélites, utilizando-se o complemento do coeficiente de coincidência simples (CS) como medida de dissimilaridade (CRUZ 1997). O CS entre dois indivíduos foi obtido dividindo o total de locos de microsatélites contendo alelos comuns pelo número total dos locos avaliados. A matriz de dados foi estabelecida atribuindo-se valor 2

para o loco de microssatélite que apresentou duas cópias do alelo homozigoto dominante, valor 1 para cada alelo quando o loco mostrou ser heterozigoto, valor 0 quando o loco apresentou duas cópias do alelo homozigoto recessivo e valor 9 quando ocorreu falha na reação.

RESULTADOS E DISCUSSÃO

O uso de marcadores microssatélites (SSR) para a caracterização de genótipos vegetais tem se tornado bastante comum em programas de melhoramento (NARVEL *et al.* 2000; TANYA *et al.* 2001; NETO *et al.* 2001). Na Figura 1 é apresentado um gel de poliacrilamida 10% onde os produtos de amplificação do *primer* Satt 181 foram aplicados. A Figura 2 mostra a distribuição dos 57 pares de *primers* de microssatélites selecionados no presente trabalho ao longo do genoma da soja. Observa-se na Figura 2 que os pares de *primers* de microssatélites selecionados estão bem distribuídos ao longo dos grupos de ligação no “mapa de ligação integrado” da soja do USDA/Iowa State University, Universidade de Utah e Universidade de Nebraska (CREGAN *et al.* 1999). Dezoito dos 24 grupos de ligação do mapa integrado da soja estão representados da seguinte forma: um grupo de ligação (A2) está representado por um SSR, cinco grupos de ligação (B1, C2, D1a, E e I) estão representados por dois SSRs, oito grupos de ligação (A1, B2/U26, D1b, F, J/U01, K, L/U14 e O) estão representados por três SSRs, três grupos de ligação (D2, H/U10 e M/U11) estão representados por quatro SSRs e um grupo de ligação (G/U05/CH11 + CH14) está representado por seis SSRs. Os grupos de ligação P (que contém um SSR mapeado), W (que contém um SSR mapeado), Q (que contém três SSR mapeados), Y (que não possui nenhum SSR mapeado), N (que contém 14 SSRs mapeados) e C1 (que contém 13 SSRs mapeados) foram, exceto o Y, amostrados, no entanto, não apresentaram SSRs polimórficos e, portanto, não estão representados no presente trabalho. Além disto, quatro dos 57 SSRs usados no presente trabalho não foram mapeados por CREGAN *et al.* (1999). Conclui-se que os grupos de ligação da soja foram bem amostrados pelos marcadores microssatélites selecionados e usados na determinação da diversidade genética dos 27 progenitores de soja. Neste sentido, HOSPITAL *et al.* (1992) observam que marcadores moleculares são muito úteis em programas de introgressão de genes, desde que estejam bem distribuídos ao longo do genoma e suas posições sejam conhecidas.

As distâncias genéticas obtidas, usando o complemento do coeficiente de coincidência simples (CS), variaram de 0,08 a 0,74 (Tabela 2),

demonstrando a grande capacidade dos marcadores microssatélites em avaliar a diversidade genética em soja. A menor distância genética encontrada (0,08) foi entre as duas linhagens CD 203PT30-1(3) e CD 203PT30-3 (4), que são doadoras de alelos para alto teor de proteína, do programa de melhoramento da qualidade da soja do BIOAGRO/UFV. Este resultado pode ser explicado pela genealogia comum destas duas linhagens uma vez que as mesmas são linhagens irmãs. A maior distância genética encontrada (0,74) foi entre a linhagem doadora de alelos para alto teor de proteína CD 203PT30-3 (4) e a variedade CD 983321 (15). Observa-se, também, na Tabela 2 que existe grande variabilidade genética entre os progenitores doadores de alelos para alto teor de proteína (0,08 a 0,70). Este resultado é importante pois indica que estes progenitores, além de representarem excelentes fontes de alelos para alto teor de proteína, apresentam grande diversidade genética, o que é desejável na escolha de progenitores para introgressão de alelos por meio de retrocruzamentos. Dentro dos progenitores recorrentes também foi observada uma grande variabilidade genética, a qual variou de 0,28 a 0,72 (Tabela 2).

A matriz de dissimilaridade foi utilizada para as análises de agrupamentos pelos métodos aglomerativos de UPGMA (“Unweighted Pair-Group Method using an Arithmetic Average”) e Vizinheiro Mais Próximo (“Single Linkage”). O dendrograma obtido pelo método Vizinheiro Mais Próximo (Figura 3) permitiu a formação de 6 grupos com as seguintes constituições: Grupo 1 formado por 14 progenitores (OC 128PT258, B1PTA25, OC 671PT100-1, OC 671PT46-1, OC 671PT46-2, CD 983321, B1PTA217, OC 13PT158-3, OC 13PT158-2, CD 985015, CD 201PT82-3, OC 14PT260, CD 211 e CD 204); Grupo 2 formado por 7 progenitores (CD 203PT30-1, CD 203PT30-3, CD 203PT30-2, CD 202PT 31-3, CD 202PT31-1, CD 202PT31-2 e CD 202PT111-1); Grupo 3 formado por 3 progenitores (OC 953194, OC 953006 e OC 953312); Grupo 4 formado por 1 progenitor (CD 983343); Grupo 5 formado por 1 progenitor (B1PTA97) e Grupo 6 formado por 1 progenitor (CD 210). O dendrograma obtido pelo método UPMGA (Figura 4) permitiu reunir os progenitores em 5 grupos distintos conforme discriminados a seguir: Grupo 1 formado por 10 progenitores (OC 128PT258, B1PTA25, OC 671PT46-1, OC 671PT46-2, CD 983321, B1PTA217, OC 14PT260, OC

671PT100-1, OC 13PT158-3, CD 210); Grupo 2 formado por 3 progenitores (CD 211, CD 204, CD 983343); Grupo 3 formado por 3 progenitores (OC 953194, OC 953006 e OC 953312); Grupo 4 formado por 4 progenitores (OC 13PT158-2, CD 985015, CD 201PT82-3 e B1PTA97); Grupo 5 formado por 7 progenitores (CD 203PT30-1, CD 203PT30-3, CD 203PT30-2, CD 202PT31-3, CD 202PT31-1, CD 202PT31-2 e CD 202PT111-1). Resultados similares foram encontrados por TANYA *et al.* (2001), que conseguiram reunir 16 variedades (genótipos) de soja em 5 grupos distintos utilizando 20 marcadores SSR pelo método UPGMA. Neste mesmo sentido, BARANEK (2002), utilizando marcadores RAPD, conseguiu reunir 19 variedades (genótipos) de soja avaliadas em cinco grupos distintos, utilizando o coeficiente de Nei e o método de agrupamento UPGMA. As linhagens doadoras que possuem origem em comum, foram reunidas, na maioria das vezes, no mesmo grupo como foi o caso das linhagens doadoras, CD 202PT e CD 203 PT, enquanto que as linhagens doadoras derivadas de OC 671 PT se reuniram em outro grupo. Todas essas linhagens doadoras foram obtidas por retrocruzamentos. O caso de algumas linhagens irmãs se reunirem em grupos diferentes, como exemplo as linhagens oriundas de B1PTA, pode ser explicado pelo reduzido número de retrocruzamentos que foram insuficientes para recuperar o genoma do progenitor recorrente. Para obter grupos mutuamente exclusivos foi utilizado o método de otimização de Tocher, que identificou 9 grupos (Tabela 3), sendo o grupo 1 formado por 7 progenitores (CD 203PT30-1, CD 203PT30-3, CD 203PT30-2, CD 202PT31-3, CD 202PT31-1, CD 202PT31-2 e CD 202PT111-1), o grupo 2 formado por 6 progenitores (OC 128PT258, OC 671PT46-1, OC 671PT46-2, CD 983321, B1PTA25 e B1PTA217); o grupo 3 formado por 3 progenitores (OC 953194, OC 953006 e OC 953312); o grupo 4 formado por 3 progenitores (OC 13PT158-2, CD 985015 e CD 201PT82-3); o grupo 5 formado por 3 progenitores (OC 14PT260, OC 671PT100-1 e OC 13PT158-3); o grupo 6 formado por 2 progenitores (CD 211 e CD 204); o grupo 7 formado por 1 progenitor (CD 983343); o grupo 8 formado por 1 progenitor (CD 210) e o grupo 9 formado por 1 progenitor (B1PTA97). Foi definido um ponto de significância para os dendrogramas do Vizinho mais Próximo de 0,40 de distância de ligação (Figura 3), e de 0,50 para UPGMA (Figura 4). Em

seguida, foram estabelecidos os grupos formados em cada dendrograma de acordo com o limite estabelecido. Uma comparação feita entre os três métodos de agrupamento (Tabela 4) demonstrou que o método de Tocher foi o que diferenciou o maior número de grupos, subdividindo o grupo 1. No entanto, os três métodos foram concordantes no agrupamento da maioria dos genótipos analisados, sendo que o Vizinho mais Próximo foi mais próximo de Tocher do que UPGMA.

Os progenitores doadores de alelos para alto teor de proteína foram selecionados com base em suas dissimilaridades genéticas, obtidas da matriz de distância genética, com os progenitores recorrentes (Tabela 5). Como pode ser visto nesta tabela, para cruzamentos com cada progenitor recorrente foram selecionados entre três a cinco progenitores doadores com distâncias genéticas variando de 0,28 a 0,57. Os números entre parênteses na Tabela 5 correspondem aos códigos dados aos genótipos que constam da Tabela 1. Assim, para o cruzamento com CD 983321 (15) foram selecionados cinco progenitores doadores (distâncias entre 0,28 a 0,48), com CD 985015 (16) quatro doadores (distâncias entre 0,31 a 0,56), com OC 953194 (18) três doadores (distâncias entre 0,49 a 0,51), com CD 211 (19) três doadores (distâncias entre 0,37 a 0,52), com CD 210 (20) quatro doadores (distâncias entre 0,43 a 0,47), com OC 953006 (21) quatro doadores (distâncias entre 0,41 a 0,47), com OC 953312 (22) quatro doadores (distâncias entre 0,41 a 0,43), com CD 983343 (23) quatro doadores (distâncias entre 0,48 a 0,54) e com CD 204 (24) quatro doadores (distâncias entre 0,42 a 0,57).

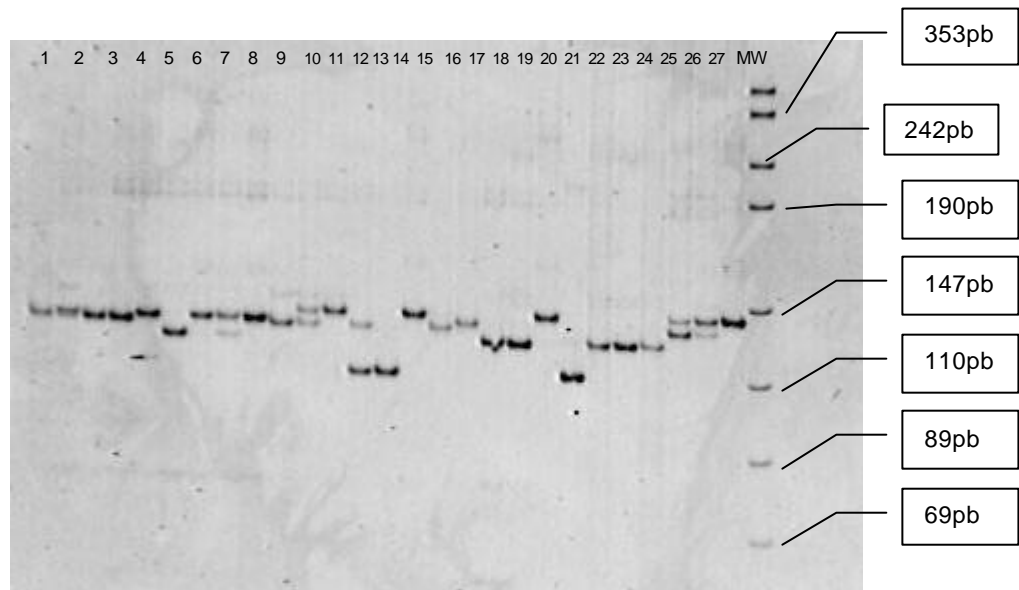
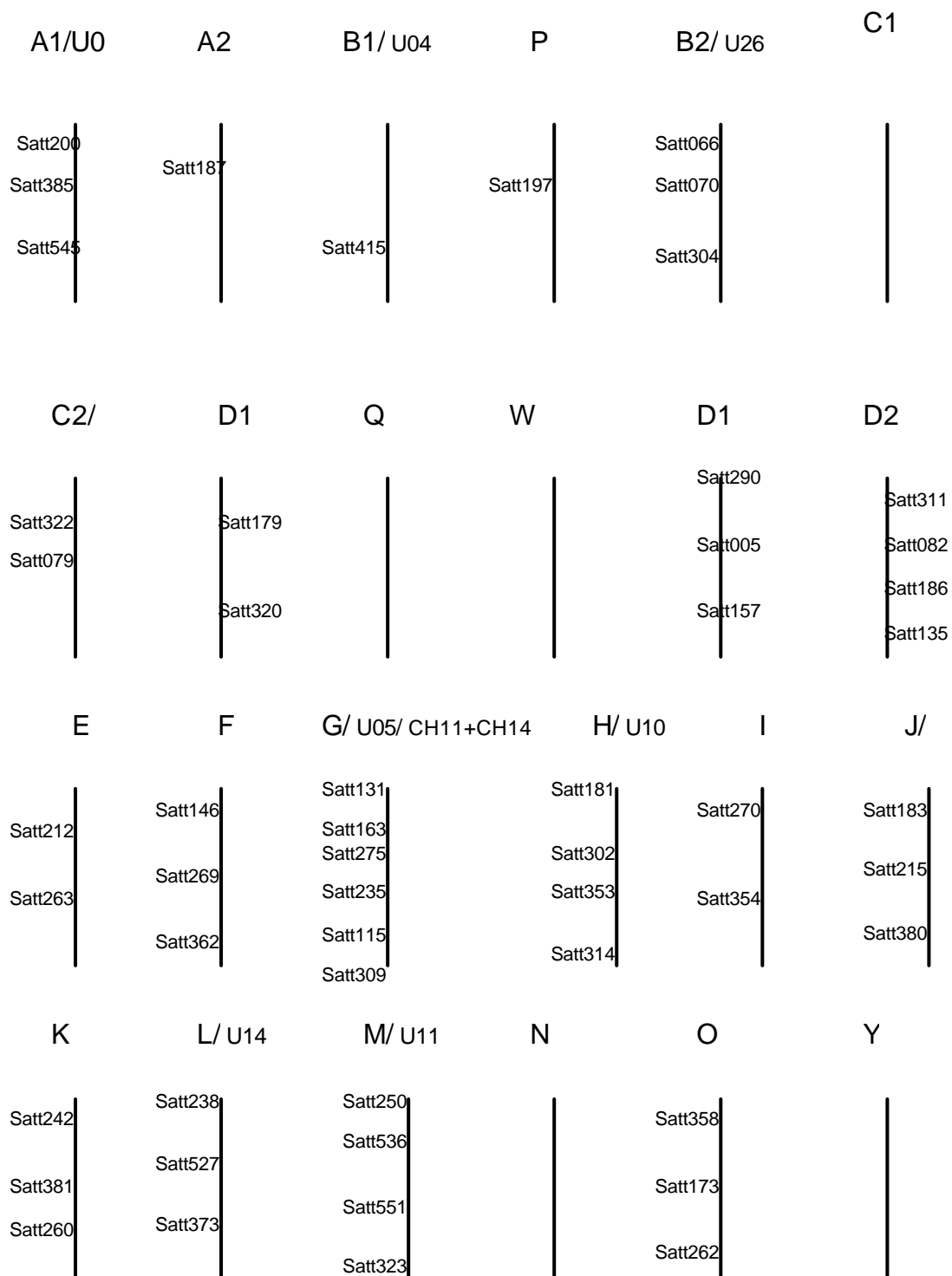


Figura 1 – Produtos de amplificação do *primer* Satt 181 nos 27 progenitores em gel de poliacrilamida 10%. A última canaleta à direita corresponde ao padrão de massa molecular do plasmídeo pUC 18 digerido com a enzima Msp I.



Não ligados: Sat009, Satt136, Satt512 e Satt574

Figura 2 – Distribuição dos 57 marcadores SSR selecionados para a análise de divergência genética no mapa de ligação integrado da soja do USDA, Universidade de Utah e Universidade de Nebraska (adaptado de Cregan *et al.* 1999).

Tabela 2 - Matriz de distâncias genéticas, calculada a partir do coeficiente de coincidência simples (CS), entre os 27 progenitores (recorrentes e doadores) utilizados no programa de retrocruzamentos para introgressão de alelos para alto teor de proteína em variedades elite de soja.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
1	0																											
2	0,5	0																										
3	0,64	0,69	0																									
4	0,63	0,66	0,08	0																								
5	0,66	0,64	0,17	0,18	0																							
6	0,61	0,65	0,33	0,38	0,39	0																						
7	0,65	0,65	0,48	0,54	0,51	0,18	0																					
8	0,68	0,7	0,34	0,38	0,37	0,13	0,15	0																				
9	0,57	0,55	0,43	0,41	0,44	0,22	0,25	0,24	0																			
10	0,53	0,41	0,55	0,54	0,6	0,54	0,5	0,57	0,42	0																		
11	0,4	0,41	0,63	0,65	0,58	0,57	0,56	0,63	0,53	0,36	0																	
12	0,39	0,37	0,65	0,67	0,59	0,58	0,58	0,65	0,53	0,32	0,1	0																
13	0,57	0,51	0,56	0,61	0,63	0,56	0,61	0,6	0,62	0,48	0,59	0,52	0															
14	0,51	0,42	0,65	0,67	0,7	0,55	0,58	0,65	0,49	0,35	0,5	0,43	0,35	0														
15	0,48	0,48	0,71	0,74	0,68	0,57	0,54	0,63	0,55	0,5	0,28	0,32	0,57	0,45	0													
16	0,6	0,56	0,56	0,62	0,62	0,55	0,57	0,6	0,64	0,56	0,56	0,56	0,31	0,58	0,65	0												
17	0,59	0,53	0,44	0,55	0,47	0,48	0,55	0,5	0,49	0,64	0,56	0,56	0,4	0,59	0,63	0,31	0											
18	0,61	0,49	0,73	0,69	0,68	0,53	0,49	0,57	0,51	0,57	0,55	0,54	0,56	0,53	0,47	0,57	0,66	0										
19	0,58	0,57	0,65	0,71	0,6	0,63	0,63	0,65	0,63	0,52	0,4	0,37	0,54	0,64	0,45	0,53	0,55	0,49	0									
20	0,58	0,49	0,68	0,72	0,69	0,63	0,61	0,68	0,61	0,43	0,52	0,46	0,47	0,44	0,5	0,48	0,57	0,48	0,52	0								
21	0,61	0,48	0,72	0,66	0,63	0,48	0,42	0,5	0,41	0,42	0,53	0,5	0,56	0,47	0,47	0,57	0,7	0,19	0,46	0,44	0							
22	0,59	0,63	0,69	0,65	0,66	0,42	0,41	0,43	0,41	0,6	0,54	0,55	0,68	0,55	0,47	0,69	0,72	0,28	0,59	0,62	0,31	0						
23	0,48	0,73	0,64	0,64	0,69	0,54	0,53	0,54	0,61	0,6	0,58	0,57	0,52	0,59	0,43	0,52	0,6	0,52	0,45	0,6	0,48	0,54	0					
24	0,49	0,63	0,65	0,64	0,67	0,67	0,72	0,66	0,68	0,62	0,57	0,54	0,59	0,6	0,55	0,63	0,58	0,69	0,39	0,56	0,73	0,6	0,42	0				
25	0,35	0,57	0,64	0,6	0,6	0,61	0,63	0,6	0,55	0,54	0,42	0,46	0,64	0,56	0,42	0,61	0,63	0,65	0,52	0,49	0,61	0,63	0,5	0,42	0			
26	0,61	0,42	0,65	0,62	0,63	0,67	0,63	0,64	0,65	0,53	0,49	0,44	0,47	0,54	0,55	0,45	0,56	0,52	0,54	0,47	0,56	0,64	0,5	0,57	0,42	0		
27	0,37	0,42	0,7	0,69	0,63	0,63	0,61	0,68	0,5	0,43	0,37	0,33	0,53	0,4	0,35	0,6	0,61	0,6	0,47	0,5	0,53	0,52	0,59	0,51	0,39	0,49	0	

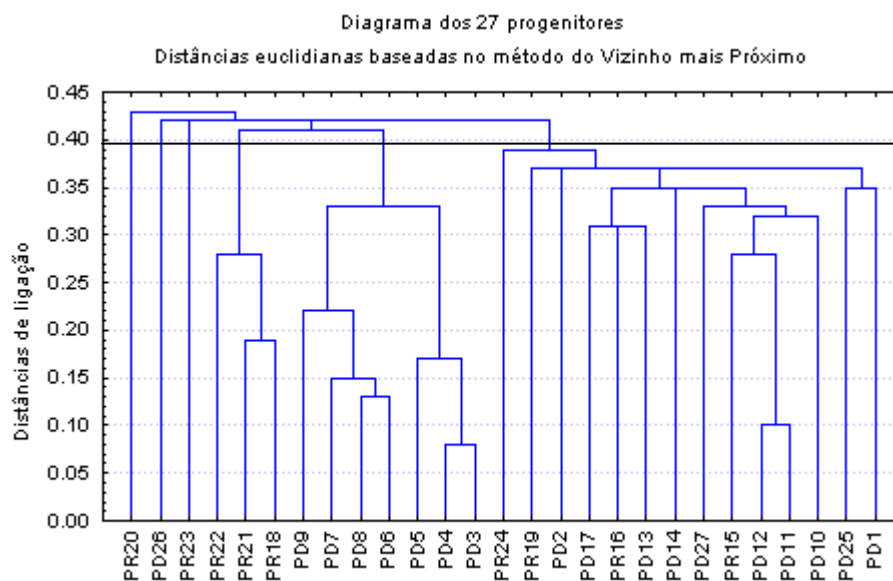


Figura 3 - Dendrograma obtido pelo método do Vizinho Mais Próximo (Single Linkage) contendo os 27 progenitores utilizados no programa de retrocruzamentos. PR indica progenitor recorrente e PD indica progenitor doador.

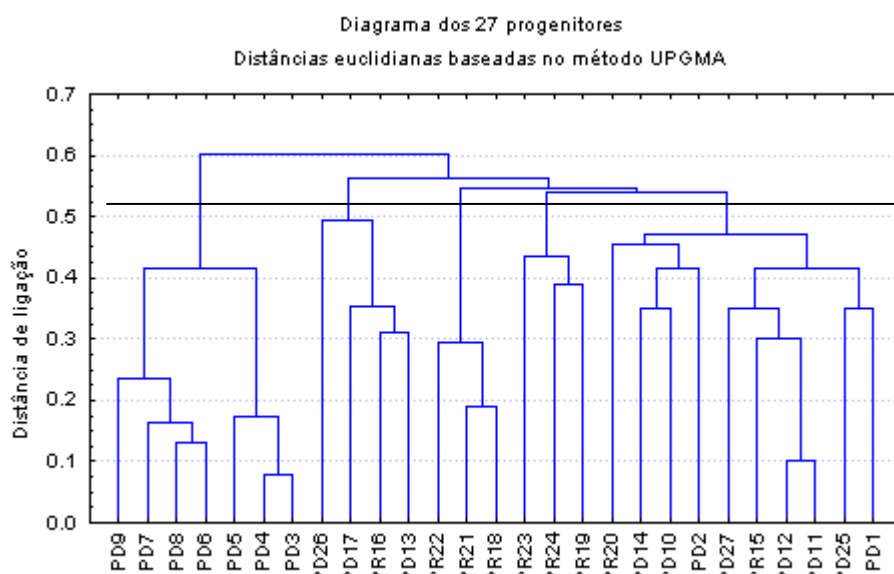


Figura 4 - Dendrograma obtido pelo método do UPGMA ("Unweighted Pair-Group Method using na Arithmetic Average") contendo os 27 progenitores utilizados no programa de retrocruzamentos. PR indica progenitor recorrente e PD indica progenitor doador.

Tabela 3 - Agrupamento obtido pelo método de Tocher dos 27 progenitores utilizados para introgressão de alelos para alto teor de proteína em variedades elite de soja. Os progenitores recorrentes estão indicados pela letra (R), os demais são progenitores doadores.

GRUPOS	PROGENITORES
1	3, 4, 5, 6, 7, 8, 9
2	1, 11, 12, 15(R), 25, 27
3	18(R), 21(R), 22(R)
4	13, 16(R), 17
5	2, 10, 14
6	19(R), 24(R)
7	23(R)
8	20(R)
9	26

Tabela 4 – Comparação entre os métodos de agrupamento. Em negrito estão os indivíduos que são comuns para os três agrupamentos.

GRUPOS	VIZINHO MAIS PRÓXIMO	UPGMA	TOCHER
1	1, 10, 11, 12, 15, 25, 27, 14, 13, 16, 17, 2, 19, 24	1, 11, 12, 15, 25, 27, 2, 10, 14, 20	3, 4, 5, 6, 7, 8, 9
2	3, 4, 5, 6, 7, 8, 7, 9	19, 24, 23	1, 11, 12, 15, 25, 27
3	18, 21, 22	18, 21, 22	18, 21, 22
4	23	13, 16, 17, 26	13, 16, 17
5	26	3, 4, 5, 6, 7, 8, 9	2, 10, 14
6	20		19, 24
7			23
8			20
9			26

Tabela 5 - Seleção dos progenitores doadores de alelos para alto teor de proteína com base em suas dissimilaridade genéticas com os progenitores recorrentes. Entre parênteses é indicado o número dado a cada progenitor constante na Tabela 1.

Progenitores recorrentes	Progenitores doadores	Distâncias genéticas
CD 983321 (15)	OC 128PT258 (1)	0,48
	OC 14PT260 (2)	0,48
	OC 671PT46-1 (11)	0,28
	OC 671PT46-2 (12)	0,32
	OC 13PT158-3 (14)	0,45
CD 985015 (16)	CD 202PT31-3 (6)	0,55
	OC 13PT158-3 (13)	0,31
	CD 203PT30-1 (3)	0,56
	CD 201PTPT82-3	0,31
OC 953194 (18)	OC 14PT260 (2)	0,49
	CD 202PT31-1 (7)	0,49
	CD 202PT111-1 (9)	0,51
CD 211 (19)	OC 671PT46-1 (11)	0,40
	OC 671PT46-2 (12)	0,37
	OC 671PT100-1 (10)	0,52
CD 210 (20)	OC 671PT100-1 (10)	0,43
	OC 13PT158-3 (14)	0,44
	OC 13PT158-2 (13)	0,47
	OC 671PT46-2 (12)	0,46
OC 953006 (21)	CD 202PT31-1 (7)	0,42
	CD 202PT111-1 (9)	0,41
	OC 671PT100-1 (10)	0,42
	OC PT13158-2 (14)	0,47
OC 953312 (22)	CD 202PT31-3 (6)	0,42
	CD 202PT31-1 (7)	0,41
	CD 202PT31-2 (8)	0,43
	CD 202PT111-1 (9)	0,41
CD 983343 (23)	OC 128PT258 (1)	0,48
	CD 202PT31-1 (7)	0,53
	CD 202PT31-1 (8)	0,54
	OC 13PT158-2 (13)	0,52
CD 204 (24)	OC 128PT258 (1)	0,49
	OC 671PT46-1 (11)	0,57
	OC 671PT46-2 (12)	0,54
	B1PTA25 (25)	0,42

REFERÊNCIAS BIBLIOGRÁFICAS

- AKKAYA, M.S., SHOEMAKER, R.C., SPECHT, J.E., BHAGWAT, A.A., CREGAN, P.B. Integration of simple sequence repeat DNA markers into a soybean linkage map. **Crop Sci.** v.35, p.1439-1445, 1995.
- BARANEK, M., KADLEC, M., RADDOVA, J., VACHUN, M., PIDRA, M. Evaluation of genetic diversity in 19 *Glycine max* (L.) Merr. accessions included in the Czech national collection of soybean genotypes. **Czech J. Genet. Plant Breed.**, v.38, n.2, p.69-74, 2002.
- BERNATZKY, R., TANKSLEY, S.D. Toward a saturated map of tomato based on isozymes and random cDNA sequences. **Genetics**, v.112, p.887-898, 1986.
- BROWN-GUEDIRA, G.L., THOMPSON, R.L., WARBURTON, M.L. Evaluation of genetic diversity of soybean introductions and north American ancestors using RAPD and SSR markers. **Crop Sci.**, v.40, p.815-823, 2000.
- CREGAN, P.B., JARVIK, T., BUSH, A.L., SHOEMAKER, R.C., LARK, K.G., KAHLER, A.L., KAYA, N., VAN, TOAI, T.T., LOHNES, D.G., CHUNG, J., SPECHT, J.E. An integrated genetic linkage map of the soybean genome. **Crop Sci.** V.39, p.1464-1490, 1999.
- CRUZ, C.D., REGAZZI, A.J. Modelos biométricos aplicados ao melhoramento genético. **Ed. UFV, Viçosa, MG**, 1997, 390p.
- CRUZ, C.D. Programa genes aplicativo computacional em genética e estatística. **Viçosa, MG: UFV**, 1997. 63p.
- HOSPITAL, F., CHEVALET, C., MULSANT, P. Using markers in gene introgression breeding programs. **Genetics**, v.132, p.1199-1210, 1992.
- MACDONALD, M.B., ELLIOT, E.J., SWEENEY, P.M. DNA extraction from seeds for RAPD analyses in varietal identification studies. **Seed Sci. & Technol.** v.22, p.171-176, 1994.
- MORGANTE, M., RAFALSKI, A., BIDDLE, P., TINGEY, S., OLIVIERI, A.M. Genetic mapping and variability of seven soybean sequence repeat loci. **Genome** v.37, p.763-769, 1994.

- NARVEL, J.M., FEHR, W.R., CHU, W.C., GRANT, D., SHOEMAKER, R.C. Simple sequence repeat diversity among soybean plant introductions and elite genotypes. **Crop Sci.** v.40, p.1452-1458, 2000.
- NETO, F.A., **Caracterização de cultivares de soja por marcadores microssatélites.** Viçosa, MG: UFV, 46p. Dissertação (Mestrado em Fitotecnia) - Universidade Federal de Viçosa, 2001.
- SAMBROOK, J., FRITSCH, E.F., MANIATIS, T., **Molecular cloning a laboratory manual.** 2nd Ed. New York, Cold Spring Harbor Laboratory Press, v.3, 1989.
- SANTOS, J.B., Emprego de marcadores moleculares no melhoramento de plantas. **Hort. Bras.** v.12, n.2, p.282-286, 1994.
- SHOEMAKER, R.C., GUFFY, R.D., LORENZEN, L.L., SPECHT, J.E. Molecular genetics mapping of soybean: Map utilization. **Crop Sci.** v.32, p.1091-1098, 1992.
- SHOEMAKER, R.C., SPECHT, J.E. Integration of the soybean molecular and classical genetic linkage groups. **Crop Sci.** v.35, p.436-446, 1995.
- TANYA, P. SRINIVES, P., TOOJINDA, T, VANAVICHIT, A., BO-KEUN HA, JEONG-SUK BAE, JUNG-KYUNG MOON, SUK-HA LEE. Evaluation of genetic diversity among soybean genotypes using SSR an SNP. **Korean J. Crop Sci.** v.46, n.4, p.334-340, 2001.

CAPÍTULO 2

ESTIMAÇÃO DO NÚMERO REFERENCIAL DE *PRIMERS* DE MICROSSATÉLITES POR SIMULAÇÃO PARA ASSISTIR UM PROGRAMA DE RETROCRUZAMENTOS EM SOJA

RESUMO

O método dos retrocruzamentos tem sido amplamente usado em programas de melhoramento para transferir características controladas por um ou pouco genes de fontes doadoras para variedades elite. A genética molecular pode ser integrada aos métodos tradicionais de seleção artificial por meio da seleção assistida por marcadores moleculares (SAM), e conseqüentemente reduzir o número de ciclos, acelerando a recuperação do genoma do progenitor recorrente em programas de retrocruzamentos. Para esta finalidade, marcadores moleculares são mais eficientes quando suas posições no mapa de ligação são conhecidas, como é o caso dos marcadores microssatélites (SSR). Definir quantos *primers* são necessários para acompanhar um programa de retrocruzamento é uma dúvida constante entre melhoristas que utilizam marcadores moleculares. Neste sentido, o objetivo deste trabalho foi o de estimar, por meio de simulações, o número e a combinação ideal de *primers* de microssatélites de um total previamente selecionado de 57, para uso na seleção assistida visando à recuperação rápida do genoma do progenitor recorrente em um programa de retrocruzamentos de introgressão de alelos para alto teor de proteína em variedades elite de soja. Para isto, três variedades de soja e seis linhagens provenientes da COODETEC com teor normal de proteína e dezoito linhagens com alto teor de proteína provenientes do BIOAGRO/UFV foram avaliadas quanto às suas diversidades genéticas, utilizando 57 pares de *primers* de microssatélites. Foi utilizado o programa GENES para número ótimo de *primers* para se proceder as simulações, definindo-se como número inicial 7 e final 57 pares de *primers*, mantendo-se constante o número de indivíduos (27) e perfazendo um total de 10 amostras a cada simulação. Com um pequeno número de *primers*, ocorre uma grande variação nos valores de correlação à medida que se aumenta o número de *primers*, esta variação é drasticamente reduzida. Com 90% de correlação o número ótimo de pares de *primers* foi de 36 e com 95% este número ótimo foi de 44 pares de *primers*. O número ótimo de *primers* escolhidos foi 44 pares de *primers* com 95% de correlação e 6,44% de estresse. Os

resultados de ordenação dos 44 *primers* pelo método de agrupamento de Tocher, mostraram que oito das 10 amostras conseguiram discriminar os 27 genótipos em nove grupos distintos, enquanto que duas amostras discriminaram os progenitores em oito grupos. Três das amostras de *primers* que discriminaram os progenitores em nove grupos mantiveram o mesmo agrupamento original dos 27 genótipos quando estes foram analisados com os 57 pares de *primers* selecionados. Conclui-se que quaisquer destas três ordenações de *primers* apresentam a mesma eficiência na discriminação dos progenitores, uma vez que os agrupamentos dos genótipos não foram alterados. Portanto, uma destas três composições de 44 pares de *primers* de SSR poderá ser escolhida e testada para iniciar um programa de retrocruzamentos contendo estes 27 progenitores visando auxiliar a recuperação rápida do genoma do progenitor recorrente.

Palavras chave: *fingerprinting*, distância genética, marcadores moleculares, seleção.

INTRODUÇÃO

O método dos retrocruzamentos tem sido utilizado para transferir genes que controlam características de herança simples para genótipos elite. Geralmente o padrão fenotípico a ser transferido é controlado por uma característica a ser transferida é controlada por um único gene. No entanto, genes de características mais complexas com altos valores de herdabilidade, também têm sido transferidos com sucesso, utilizando o método dos retrocruzamentos (OPENSHAW *et al.* 1994). Diversos livros textos de melhoramento de plantas provêm boas descrições do procedimento de retrocruzamento (ALLARD 1960; FEHR 1987). Um parental doador (PD) que carrega a característica de interesse é cruzado com um progenitor recorrente (PR), o qual é uma variedade elite que não possui a característica de interesse. A geração F1 é cruzada com o PR para produzir a geração RC1. Na geração RC1, e nas gerações subseqüentes, os indivíduos que carregam o gene que está sendo transferido são retrocruzados com o PR. A proporção esperada do genoma do PD é reduzida pela metade em cada geração de retrocruzamento. Ignorando o efeito de ligação do alelo selecionado a ser transferido, a porcentagem do genoma do progenitor recorrente (%PR) esperada em cada geração de retrocruzamento pode ser calculada da seguinte maneira:

$$\%PR = 100 [1 - (0,5)^{n+1}]$$

em que:

n = nº de retrocruzamentos.

Após seis gerações de retrocruzamentos, a proporção esperada de recuperação do genoma do PR é >99% (Tabela 1). Até recentemente, as discussões sobre a recuperação do genoma do PR durante as gerações de retrocruzamentos têm enfatizado os valores esperados para %PR como mostrados na Tabela 1, não levando em consideração a variação da %PR que existe em torno da média do valor esperado (OPENSHAW *et al.* 1994).

Tabela 1 – Porcentagem de recuperação do progenitor recorrente durante as gerações de retrocruzamentos, assumindo que não existe ligação ao gene a ser transferido

Geração	% PR
F1	50,0000
RC1	75,0000
RC2	87,5000
RC3	93,7500
RC4	96,8750
RC5	98,4375
RC6	99,2188
RC7	99,6094

Com o desenvolvimento de marcadores genéticos capazes de promover uma boa cobertura do genoma, tem crescido o interesse de se aproveitar esta variação para aumentar a eficiência do método de retrocruzamento. Marcadores moleculares são fáceis de serem usados, pois já existem em grande número para várias espécies vegetais de interesse. Locus marcadores que são neutros e fáceis de serem identificados oferecem a oportunidade de serem usados para calcular diretamente a contribuição de cada progenitor (VANTOAI *et al.* 1997; LORENZEN *et al.* 1995). Neste sentido, em cada geração do programa de melhoramento, os genótipos que carregam o gene a ser introduzidoB podem ser selecionados e ainda, entre estes, aqueles que carregam a menor proporção do genoma do progenitor doador em outros locos diferentes daqueles que contém o gene de interesse (HOSPITAL *et al.* 1992). Embora seja relativamente fácil de montar um experimento de seleção assistida por marcadores moleculares, a escolha da estratégia mais apropriada e eficiente não é tão direta devido ao número de variáveis envolvidas. Antes de montar qualquer experimento o número de genes envolvidos na seleção e o nível de recuperação do genoma devem ser definidos. Na seleção assistida por marcadores moleculares, o nível de recuperação do genoma está fortemente associado ao número e a distribuição dos marcadores de DNA em locos que não são alvos e as

freqüências de recombinação entre o gene alvo e os marcadores que o flanqueiam (RIBAUT *et al.* 2002). OPENSHAW *et al.* (1994), em um estudo de simulação, relatam que cerca de 80 marcadores codominantes uniformemente distribuídos ao longo do genoma do milho são suficientes para recuperar 99% do genoma do progenitor recorrente em três gerações de retrocruzamentos, enquanto que sem o uso destes marcadores esta mesma estimativa somente é alcançada com seis a sete gerações de retrocruzamentos. Este autor chama a atenção para a importância do número de marcadores a serem usados, pois, quando este número é reduzido há tendência de superestimar a recuperação do genoma do progenitor recorrente. Com o uso de poucos marcadores eles se tornam fixados rapidamente, enquanto que o restante do genoma ainda não está fixado. OPENSHAW *et al.* (1994) sugerem o uso de quatro marcadores para cada cromossomo de 200 cM (caso do milho) e VISSCHER *et al.* (1996), sugere ainda, marcadores a cada 10 a 20 cM na região que flanqueia o gene de interesse. Segundo OPENSHAW *et al.* (1994), com o uso de marcadores moleculares, a porção do genoma recorrente que não está ligada ao alelo que está sendo transferido é recuperada rapidamente e com alta confiabilidade.

A inter-relação entre indivíduos que estão sendo analisados pode ser melhor visualizada em um gráfico bidimensional, desde que este represente com fidelidade o nível de associação entre os indivíduos no espaço n-dimensional. A diminuição do espaço n-dimensional para um espaço bidimensional pode ser obtida pelo uso de componentes principais, analisando-se os valores originais das variáveis (CRUZ 1997). Segundo CRUZ e VIANA (1994), a eficiência da transformação do espaço n-dimensional para o espaço bidimensional pode ser avaliada pelas correlações entre as distâncias estimadas baseadas nas coordenadas obtidas e nas distâncias originais; pelo grau de distorção das distâncias devido à projeção no espaço bidimensional; e pelo coeficiente de estresse. A correlação entre as distâncias originais e as estimadas é obtida da mesma forma que a correlação cofenética. Quando o estresse produzido pela conversão do espaço n-dimensional para o espaço bidimensional for menor

que 20%, pode-se então utilizar a dispersão gráfica para avaliar o inter-relacionamento entre os indivíduos avaliados. À medida que o estresse aumenta acima de 20% tende-se a cometer erros, pois a situação avaliada pode não mais corresponder à situação real.

O objetivo principal deste trabalho foi estimar, por meio de simulações, o número e a combinação ideal de *primers* de microssatélites de um total previamente selecionado de 57, para uso na seleção assistida visando à recuperação rápida do genoma do progenitor recorrente em um programa de retrocruzamentos de introgressão de alelos para alto teor de proteína em variedades elite de soja.

MATERIAL E MÉTODOS

Material genético

Três variedades (CD 204, CD 210, CD 211) e seis linhagens de soja (CD 983321, CD 985015, CD 983343, OC 953006, OC 953194, e OC 953312) provenientes da Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda - COODETEC constantes da Tabela 1 foram utilizadas como progenitores recorrentes. Dezoito linhagens, com alto teor de proteína do programa de melhoramento da soja do BIOAGRO/UFV, foram utilizadas como fontes doadoras da característica alto teor de proteína com o objetivo de se incorporar alelos para alto teor de proteína nas variedades da COODETEC.

Tabela 2 - Progenitores usados no programa de retrocruzamentos visando a incorporação de alelos para alto teor de proteína em variedades de soja da COODETEC.

Progenitor	Genótipo	Categoria
1	OC 128PT258	Doador
2	OC 14PT260	Doador
3	CD 203PT30-1	Doador
4	CD 203PT30-3	Doador
5	CD 203PT30-2	Doador
6	CD 202PT31-3	Doador
7	CD 202PT31-1	Doador
8	CD 202PT31-2	Doador
9	CD 202PT111-1	Doador
10	OC 671PT100-1	Doador
11	OC 671PT46-1	Doador
12	OC 671PT46-2	Doador
13	OC 13PT158-2	Doador
14	OC 13PT158-3	Doador
15	CD 983321	Recorrente
16	CD 985015	Recorrente
17	CD 201PT82-3	Doador
18	OC 953194	Recorrente
19	CD 211	Recorrente
20	CD 210	Recorrente
21	OC 953006	Recorrente
22	OC 953312	Recorrente
23	CD 983343	Recorrente
24	CD 204	Recorrente
25	B1PTA25	Doador
26	B1PTA97	Doador
27	B1PTA217	Doador

Extração e purificação de DNA

Amostras de DNA das sementes foram extraídas pelo método descrito segundo MCDONALD *et al.* (1994) com algumas modificações. Em tubo eppendorff de 1,5 mL contendo cerca de 50 mg de sementes eram adicionados 1000 µL de tampão de extração contendo Tris-HCl 0,2 M, NaCl 0,28 M, EDTA 25 mM e SDS 10%. As amostras eram extraídas em Politron e, em seguida, centrifugadas por 10 min por 14.000 RPM em microcentrífuga eppendorff 1405. Os sobrenadantes eram transferidos para novos tubos, sendo, em seguida, acrescentados 10 µL de proteinase K 10mg/mL e 10 µL de CaCl₂ 1 mM, e colocado em banho-maria a 55 °C por 30 min. Eram, então, adicionados às amostras 900 µL de isopropanol e deixadas em repouso por 2 min. Após este tempo as amostras eram centrifugadas por 10 min a 14.000 RPM. Os sobrenadantes eram descartados e os precipitados secos por 15 min à temperatura ambiente sendo, a seguir, ressuspensos em solução de Tris-EDTA (TE) 10 mM e 1 mM contendo RNase 60 µg/mL e colocados em banho-maria por 90 min. As amostras eram precipitadas novamente pela adição de 900 µL de isopropanol e deixadas em precipitação por 2 min. Logo após as amostras eram novamente centrifugadas por 10 min a 14.000 RPM e os sobrenadantes eram descartados. Os precipitados eram ressuspensos ao final em TE 10 mM e 1 mM. A concentração do DNA era estimada espectrofotometricamente por leitura da absorbância em 260 nm, sendo cada unidade de absorbância correspondente a 50 µg/mL de DNA fita dupla (SAMBROOK *et al.* 1989).

Reações de amplificação

As reações de amplificação de microssatélites eram feitas em um volume total de 15 µL, contendo 12,5 µM de Tris-HCl (pH 8,3), 62,5 µM de KCl, 2,5 mM de MgCl₂, 125 µM de cada um dos deoxinucleotídios (dATP, dTTP, dGTP e dCTP), 0,6 µM de cada *primer*, uma unidade da enzima Taq-polimerase e 18 ng de DNA. As amplificações eram realizadas em Termociclador Perkin Elmer 9600 programado para uma etapa inicial de 7

min. a 72 °C, seguida de 30 ciclos de 1 min. a 94 °C, 1 min. a 50 °C e 2 min. a 72 °C. Por fim, uma etapa de 7 min. a 72 °C. Os fragmentos amplificados eram separados por eletroforese em gel não desnaturante de poliacrilamida 10% por 4 horas a 100 V, e corados com solução de brometo de etídio (0,2 µg/mL). O gel possuía 16,5x12,5 cm e espessura de 0,5 mm. Terminada a eletroforese, os géis eram fotodigitalizados pelo sistema de fotodocumentação Eagle Eye II (Stratagene).

Seleção de pares de primers de microssatélites

De um total de cerca de 600 pares de *primers* de microssatélites disponíveis para soja, foram testados 86 e selecionados 57 com base no mapa integrado da soja (CREGAN *et al.* 1999), os quais se mostraram polimórficos para obter os dados moleculares dos genótipos, sendo eles listados a seguir: (1)Sat009, (2)Satt005, (3)Satt066, (4)Satt070, (5)Satt079, (6)Satt082, (7)Satt115, (8)Satt131, (9)Satt135, (10)Satt136, (11)Satt146, (12)Satt157, (13)Satt163, (14)Satt173, (15)Satt179, (16)Satt181, (17)Satt183, (18)Satt186, (19)Satt187, (20)Satt197, (21)Satt200, (22)Satt212, (23)Satt215, (24)Satt235, (25)Satt238, (26)Satt242, (27)Satt250, (28)Satt260, (29)Satt262, (30)Satt263, (31)Satt269, (32)Satt270, (33)Satt275, (34)Satt290, (35)Satt304, (36)Satt302, (37)Satt309, (38)Satt311, (39)Satt314, (40)Satt320, (41)Satt322, (42)Satt323, (43)Satt353, (44)Satt358, (45)Satt354, (46)Satt362, (47)Satt373, (48)Satt380, (49)Satt381, (50)Satt385, (51)Satt415, (52)Satt512, (53)Satt527, (54)Satt536, (55)Satt545, (56)Satt551 e (57)Satt574.

Estimadores das distâncias genéticas

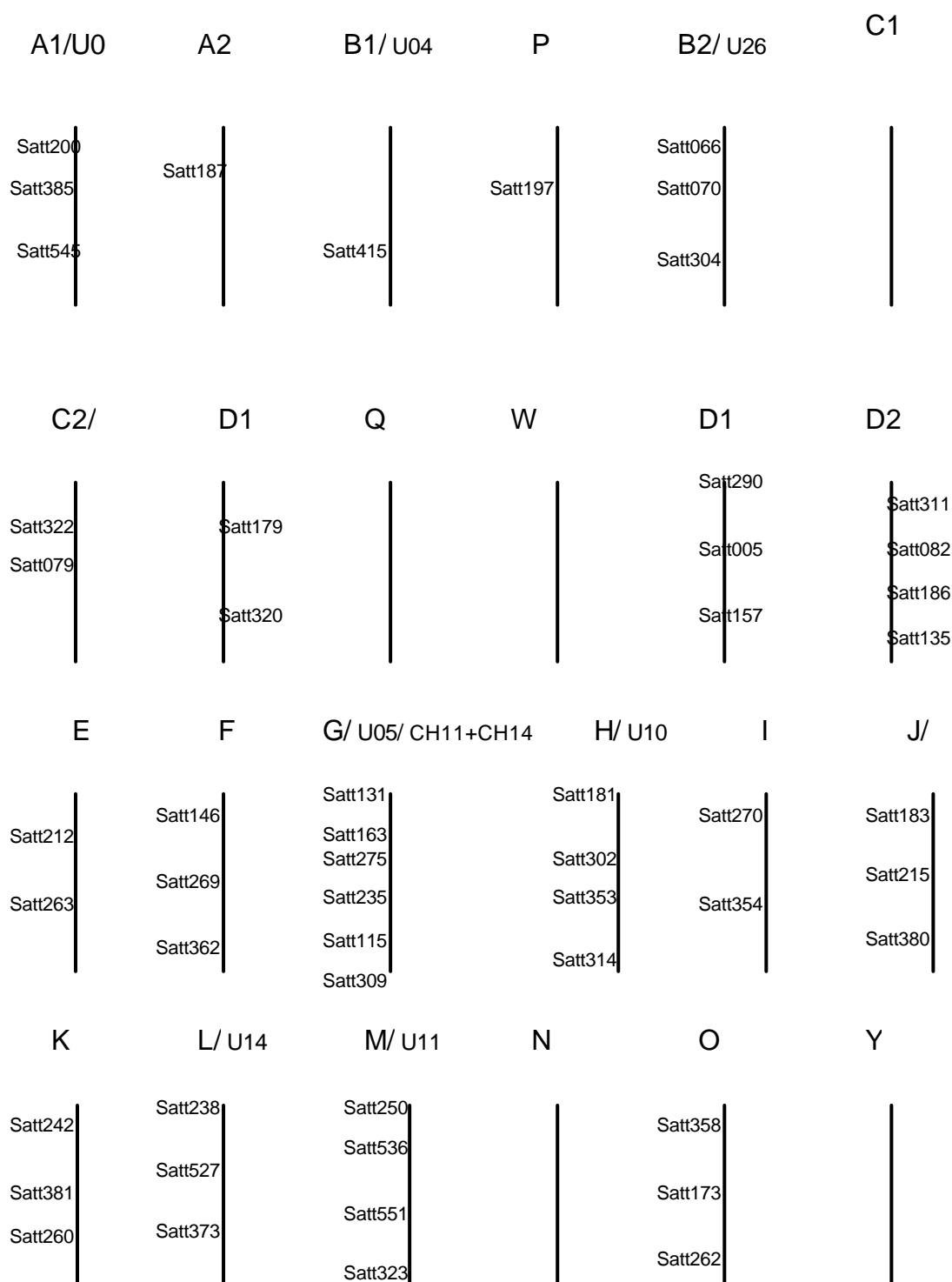
As distâncias genéticas entre os progenitores foram estimadas com base nos dados das análises feitas com os microssatélites, utilizando-se o complemento do coeficiente de coincidência simples (CS) como medida de dissimilaridade (CRUZ 1997). O CS entre dois indivíduos foi obtido dividindo

o total de locos de microssatélites contendo alelos comuns pelo número total dos locos avaliados. A matriz de dados foi estabelecida atribuindo-se valor 2 para o loco de microssatélite que apresentou duas cópias do alelo homocigoto dominante, valor 1 para cada alelo quando o loco mostrou ser heterocigoto, valor 0 quando o loco apresentou duas cópias do alelo homocigoto recessivo e valor 9 quando ocorreu falha na reação.

Simulação

Esta simulação foi feita baseada no número final de pares de *primers* previamente selecionados (57), assumindo que este seria um número suficiente para se executar este estudo de simulação, uma vez que os mesmos estão bem distribuídos ao longo do genoma (Figura 1), o que está de acordo as observações de HOSPITAL *et al.* (1992). Tomando-se, então, 57 pares de *primers* de SSR não aleatórios e bem distribuídos para discriminar os progenitores, questionou-se, portanto, qual seria o número de *primers* ideal para acompanhar a recuperação do genoma do progenitor recorrente em um programa de retrocruzamentos. O número inicial estabelecido para a simulação foi o de 7 pares de *primers*, mantendo-se constante o número de indivíduos (27) e tomando-se ao acaso 10 amostras por simulação, com um incremento de um marcador a cada simulação. Para fins de processamento foi obtida uma matriz original de dissimilaridade de dimensão 27 x 27 e 500 outras matrizes de mesma dimensão, obtidas a partir de 7 a 56 *primers*, repetidos 10 vezes, amostrados por *bootstrap*. A medida de eficiência adotada neste trabalho foi a de correlação. As estimativas de correlação escolhidas foram de 90% e 95%. Uma outra medida de eficiência que pode ser adotada é o estresse que também foi calculado. Uma outra análise de simulação efetuada pelo método de agrupamento de Tocher foi usada para determinar, por meio do número ótimo de *primers* determinado na simulação anterior: (a) se quaisquer das combinações de *primers* selecionados mostram o mesmo perfil de agrupamento e, (b) quais são as combinações de *primers* que melhor se encaixam neste perfil de agrupamento. É importante salientar que estes

estudos de simulação se basearam em dados que foram produzidos a partir de análises que foram realizadas. Muitos dos estudos de simulação que são encontrados na literatura estão baseados em pressuposições e não em dados reais.



Não ligados: Sat009, Satt136, Satt512 e Satt574

Figura 2 – Distribuição dos 57 marcadores SSR selecionados para a análise de divergência genética no mapa de ligação integrado da soja do USDA, Universidade de Utah e Universidade de Nebraska (adaptado de Cregan *et al.* 1999).

RESULTADOS E DISCUSSÃO

A Tabela 3 apresenta os resultados da simulação para 7 e 44 *primers* a título de ilustração, a qual se iniciou com 7 e terminou com 57 *primers*, mantendo-se constante o número de genótipos (27). Os valores de correlação, soma de quadrado de desvio e de estresse para todas as simulações realizadas estão contidas na Tabela 1 do Anexo. Para isto, foram tomadas ao acaso 10 amostras, com um incremento de um *primer* a cada simulação até atingir os 57 *primers* finais. Como se observa na Tabela 3, com um pequeno número de *primers*, ocorre uma grande variação nos valores de correlação, soma de quadrado de desvio e do estresse, ou seja, com 7 *primers* os valores de correlação variaram de 0,4071 a 0,7015. À medida que se aumenta o número de *primers*, esta variação é drasticamente reduzida. Como mostrado na Tabela 3, quando são utilizados 44 *primers*, os valores de correlação variam muito poucos, ou seja, de 0,9540 a 0,9696.

Uma maneira mais direta de observar esta simulação é por meio da análise gráfica desses dados (Figuras 2 e 3). Com correlação de 0,90, entre as medidas de dissimilaridades obtidas com 57 e as obtidas com menor número de *primers*, o número ótimo de pares de *primers* foi de 36 e com correlação de 0,95 este número ótimo foi de 44 pares de *primers* (Figura 2). A soja (*Glycine max*) possui número de cromossomos $2n = 40$ (PALMER *et al.* 1996). Tomando-se 44 pares de *primers*, como número ótimo, distribuídos uniformemente ao longo do genoma, isto resultaria em 2,2 pares de *primers* por cromossomo no genoma da soja. OPENSHAW *et al.* (1994), investigando o número de marcadores necessários para aumentar a eficiência da recuperação do genoma do progenitor recorrente num programa de introgressão de genes, a partir de estudos de simulação para o genoma do milho, concluíram que seriam necessários cerca de quatro marcadores codominantes para cada cromossomo de 200 cM, na geração RC1. Por outro lado, HOSPITAL *et al.* (1992) relatam que a utilização de um número de marcadores maior do que três por cromossomo não carregador (cromossomo que não contém o gene alvo) não foi eficiente em gerações iniciais de retrocruzamento.

Ainda na Figura 2, pode ser observado que a utilização de pequeno número de marcadores, causou grande dispersão (variação) nos valores de correlação das amostras e acarretou altos valores de estresses, em torno de 30% (Tabela 3). À medida que o número de marcadores aumenta, os valores de correlação vão se tornando mais compactos e uniformes, o que acontece por volta do número de marcadores igual a 30, onde o estresse reduz consideravelmente (cerca de 10% na Tabela 1 do Anexo). Segundo CRUZ e VIANA (1994), correlação acima de 0,80 aliada a um estresse abaixo de 20%, são bons indicativos para uma análise gráfica quando se tem um número muito grande de genótipos para ser avaliado. Considerando o estresse provocado pela simulação do número de marcadores na Figura 3, pode-se concluir que para um estresse de 10% o número de marcadores ótimo foi de 41 e para um estresse de 5% o número ótimo de marcadores foi de 52. Novamente pode ser observado na Figura 3 que à medida que se aumenta o número de marcadores a variação nos valores de estresse vai sendo reduzida. Foi escolhido então, 44 pares de *primers*, como sendo um número ótimo, o qual apresenta 95% de correlação, com um valor médio de estresse de 6,44%.

A seguir, simulou-se o efeito da composição dos 44 pares de *primers* de SSR, considerados como número ótimo para assistir a recuperação do genoma do progenitor recorrente, em um total de 57 avaliados. Esta próxima simulação, baseada no método de agrupamento de Tocher, definiu quais foram os *primers* que melhor discriminaram os genótipos. A simulação foi efetuada tomando-se 44 *primers*, dos 57 avaliados, sorteados ao acaso em um total de 10 amostras. O resultado destas análises mostrou que oito das 10 amostras conseguiram discriminar os 27 genótipos em nove grupos distintos, enquanto que duas amostras discriminaram os progenitores em oito grupos (Tabelas 2 a 11 do Anexo). Três das amostras de *primers* que discriminaram os progenitores em nove grupos (Tabela 5) mantiveram o mesmo agrupamento original dos 27 genótipos quando estes foram analisados com os 57 pares de *primers* selecionados (Tabela 4). Conclui-se que quaisquer destas três ordenações de *primers* apresentam a mesma eficiência na discriminação dos progenitores, uma vez que os agrupamentos

dos genótipos não foram alterados. Na tabela 6 é apresentada a ordem dos 44 pares de *primers* que foram utilizadas nas amostras 4, 5 e 9. Portanto, uma destas três composições de 44 pares de *primers* de SSR poderá ser escolhida e testada para iniciar um programa de retrocruzamento contendo estes 27 progenitores visando auxiliar a recuperação rápida do genoma do progenitor recorrente. É importante ressaltar que, conforme mostrado em simulações realizadas por HOSPITAL *et al.* (1992), a fixação dos marcadores moleculares ocorre muito rapidamente, quando se emprega uma alta intensidade de seleção, num programa de introgressão de genes e que, após esta fixação, estes marcadores se tornam inúteis.

HOSPITAL *et al.* (1992) mostraram, por meio de simulações, que a seleção assistida por marcadores moleculares pode levar a um ganho de tempo, de no mínimo, duas gerações; e que, ao contrário do que é comumente recomendada, a seleção é mais eficiente quando realizada em gerações precoces para marcadores distantes do gene a ser introduzido e, nas gerações mais tardias, para marcadores próximos ao gene. Os autores mostraram ainda que a seleção preliminar no segmento do gene a ser inserido pode reduzir a intensidade de seleção no restante do genoma.

Resultados de simulações para diferentes estratégias, usando o genoma do milho como modelo, para comparar a introgressão de alelos assistida por marcadores moleculares através de programa de retrocruzamentos, indicam que a resposta à seleção no RC1 poderia ser aumentada significativamente quando o tamanho da população selecionada for inferior a 50, e que ocorre uma diminuição na resposta de seleção quando este tamanho for maior que 100 (RIBAUT *et al.* 2002). Também foi demonstrado que a seleção assistida por marcadores moleculares (SAM) em gerações mais avançadas de retrocruzamentos (RC3 comparada com RC2 e RC1) é mais eficiente pois a proporção do desvio padrão em relação à média da contribuição do genoma do doador aumenta à medida que ocorrem as sucessivas gerações de retrocruzamentos.

Em todos os casos, estes autores reforçam a necessidade de se colocar esforços iniciais na identificação dos mais convenientes conjuntos de marcadores moleculares a serem usados nas diferentes estratégias de

retrocruzamentos assistidos por marcadores moleculares (RC-SAM).

Tabela 3 – Efeito da simulação sobre o número inicial de 7 pares de *primers* e o número ideal de 44 pares de *primers*.

Número de <i>primers</i>	Amostra	Correlação	SQDesvio	Estresse
7	1	0,5499	10,8213	0,3196
7	2	0,5947	8,1114	0,2767
7	3	0,4974	10,8868	0,3206
7	4	0,4399	10,7834	0,3191
7	5	0,7015	10,8143	0,3195
7	6	0,4703	9,2412	0,2954
7	7	0,4071	9,6423	0,3017
7	8	0,482	11,6505	0,3316
7	9	0,6518	8,4837	0,283
7	10	0,4787	8,8219	0,2886
7	Média	0,5273	9,9257	0,3056
44	1	0,954	0,54	0,0714
44	2	0,9548	0,6554	0,0787
44	3	0,9694	0,3623	0,0585
44	4	0,9695	0,3746	0,0595
44	5	0,967	0,5341	0,071
44	6	0,9631	0,4482	0,0651
44	7	0,9696	0,398	0,0613
44	8	0,9661	0,3428	0,0569
44	9	0,9591	0,4553	0,0656
44	10	0,9671	0,3394	0,0566
44	Média	0,964	0,445	0,0644

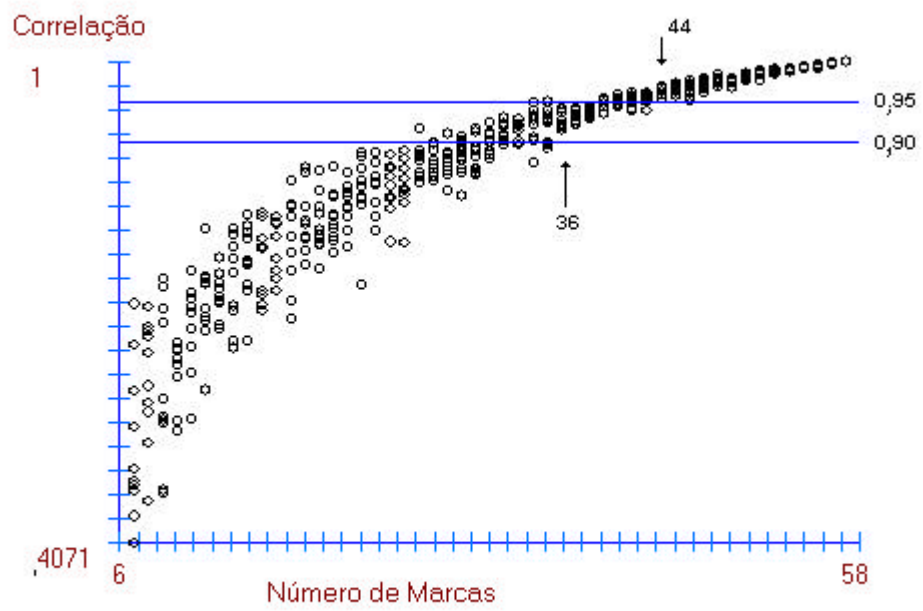


Figura 2 – Simulação do número ótimo de marcadores SSR assumindo valores de correlação de 0,90 e 0,95 e considerando o número final de 57 pares de *primers* de SSR selecionados.

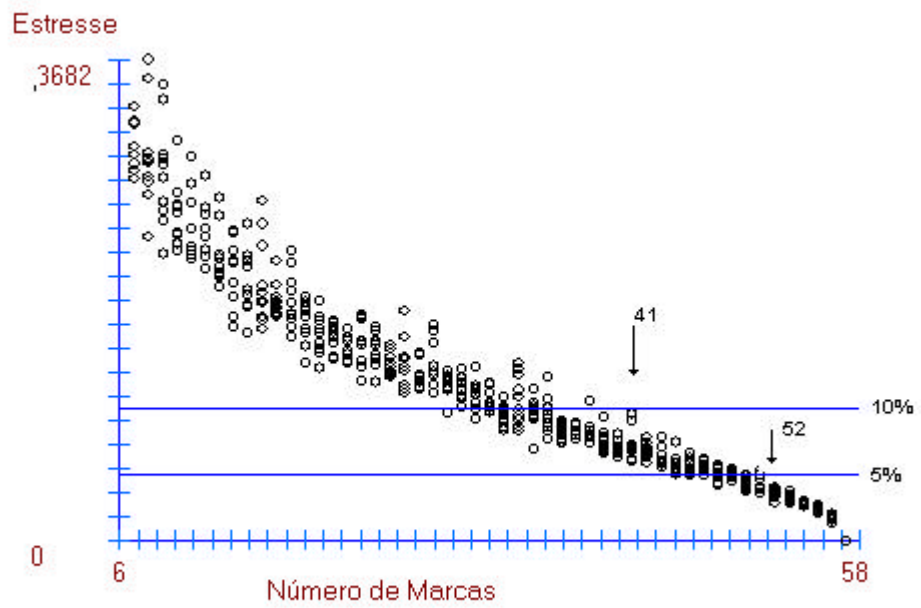


Figura 3 – Simulação do número ótimo de marcadores SSR assumindo valores de estresse de 5% e 10% e considerando um total de 57 pares de *primers* de SSR previamente selecionados.

Tabela 3 - Agrupamento obtido pelo método de Tocher dos 27 progenitores utilizados para introgressão de alelos para alto teor de proteína em variedades elite de soja. Os progenitores recorrentes estão indicados pela letra (R), os demais são progenitores doadores.

GRUPOS	PROGENITORES
1	3, 4, 5, 6, 7, 8, 9
2	1, 11, 12, 15(R), 25, 27
3	18(R), 21(R), 22(R)
4	13, 16(R), 17
5	2, 10, 14
6	19(R), 24(R)
7	23(R)
8	20(R)
9	26

Tabela 5 – Simulação dos agrupamentos com o número ótimo de 44 *primers* para as amostras aleatórias 4, 5 e 9.

Grupo	Genótipos
1	3 4 5 8 6 9 7
2	11 12 27 15 10 14 2
3	18 21 22
4	16 17 13
5	19 24
6	1 25
7	23
8	26
9	20

Tabela 6 – Ordem dos 44 pares de *primers* tomados ao acaso nas amostras 4, 5 e 9 que mantêm o mesmo padrão de agrupamento dos 57 *primers* SSR original.

Amostras	Ordem dos 44 pares de <i>primers</i>
4	(1) Sat009, (4) Satt070, (5) Satt079, (6) Satt082, (7) Satt115, (8) Satt131, (9) Satt135, (10) Satt136, (11) Satt146, (13) Satt163, (14) Satt173, (17) Satt183, (18) Satt186, (20) Satt197, (21) Satt200, (22) Satt212, (23) Satt215, (24) Satt235, (26) Satt242, (27) Satt250, (28) Satt260, (29) Satt262, (30) Satt263, (31) Satt269, (32) Satt270, (33) Satt275, (34) Satt290, (36) Satt302, (37) Satt309, (38) Satt311, (39) Satt314, (40) Satt320, (41) Satt322, (42) Satt323, (43) Satt353, (44) Satt358, (46) Satt362, (47) Satt373, (48) Satt380, (50) Satt385, (51) Satt415, (52) Satt512, (55) Satt545, e (56) Satt551.
5	(1) Sat009, (2) Satt005, (3) Satt066, (4) Satt070, (5) Satt079, (6) Satt082, (7) Satt115, (9) Satt135, (10) Satt136, (11) Satt146, (12) Satt157, (13) Satt163, (14) Satt173, (15) Satt179, (16) Satt181, (18) Satt186, (20) Satt197, (21) Satt200, (22) Satt212, (23) Satt215, (24) Satt235, (25) Satt238, (27) Satt250, (28) Satt260, (29) Satt262, (30) Satt263, (32) Satt270, (34) Satt290, (35) Satt304, (36) Satt302, (37) Satt309, (39) Satt314, (40) Satt320, (42) Satt323, (44) Satt358, (45) Satt354, (47) Satt373, (48) Satt380, (49) Satt381, (50) Satt385, (51) Satt415, (54) Satt536, (56) Satt551 e (57) Satt574.
9	(1) Sat009, (3) Satt066, (4) Satt070, (5) Satt079, (6) Satt082, (10) Satt136, (11) Satt146, (12) Satt157, (16) Satt181, (20) Satt197, (21) Satt200, (22) Satt212, (23) Satt215, (24) Satt235, (25) Satt238, (26) Satt242, (27) Satt250, (28) Satt260, (29) Satt262, (30) Satt263, (31) Satt269, (32) Satt270, (33) Satt275, (34) Satt290, (36) Satt302, (37) Satt309, (38) Satt311, (39) Satt314, (40) Satt320, (41) Satt322, (43) Satt353, (44) Satt358, (45) Satt354, (46) Satt362, (47) Satt373, (48) Satt380, (49) Satt381, (50) Satt385, (51) Satt415, (52) Satt512, (54) Satt536, (55) Satt545, (56) Satt551 e (57) Satt574.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALLARD, R.W. Principles of plant breeding. **Wiley**, New York, 1960.
- CRUZ, C.D. **Programa genes aplicativo computacional em genética e estatística**. Viçosa, MG: UFV, 1997. 63p.
- CREGAN, P.B., JARVIK. T., BUSH, A.L., SHOEMAKER, R.C., LARK, K.G., KAHLER, A.L., KAYA, N., VAN, TOAI, T.T., LOHNES, D.G., CHUNG, J., SPECHT, J.E. An integrated genetic linkage map of the soybean genome. **Crop Sci.** V.39, p.1464-1490, 1999.
- CRUZ, C.D. VIANA, J.M.S. A methodology of genetic divergence analysis based on sample unit projection on two-dimensional space. on two-dimensional space. **Rev. Brazil. Genet.**, v.17, n.1, p.69-73, 1994.
- FEHR, W.F. Principles of cultivar development, v.1, Theory and Technique, **Macmillan**, New York, 1987.
- HOSPITAL, F., CHEVALET, C., MULSANT, P. Using markers in gene introgression breeding programs. **Genetics**, v.132, p.1199-1210, 1992.
- LORENZEN, L.L., BOUTIN. S., YOUNG, N., SPECHT, E., SHOEMAKER, R.C. Soybean pedigree analysis using map-based molecular markers: I. Tracking RFLP markers in cultivars. **Crop Sci.**, v.35, p.1326-1336, 1995.
- MAcDONALD, M.B., ELLIOT, E.J., SWEENEY, P.M. DNA extraction from seeds for RAPD analyses in varietal identification studies. **Seed Sci. & Technol.** v.22, p.171-176, 1994.
- OPENSHAW, S.J.; JARBOE, S.G. e BEAVIS, W.D. Marker-assisted selection in backcross breeding. In: ASHS/CSSA Joint **Plant Breeding Symposium, 2, Corvallis Proceedings...** Corvallis: Oregon State University, 1994.
- PALMER, R.G., HYMOWITZ, T. NELSON, R.L. Germoplasm diversity within soybean. In: **Soybean genetics, molecular biology and biotechnology**. Ohio, USA:CAB INTERNACIONAL, p.1-35, 1996.
- RIBAUT, J.M., JIANG, C., HOISINGTON, D. Simulation experiments on efficiencies of gene introgression by backcross. **Crop Sci.**, v.42, p.557-565, 2002.

SAMBROOK, J., FRITSCH, E.F., MANIATIS, T., **Molecular cloning a laboratory manual**. 2^o Ed. New York, Cold Spring Harbor Laboratory Press, v.3, 1989.

VANTOAI, T.T., PENG, J., ST. MARTIN, S.K. Using AFLP to determine the genomic contribution of parents to populations. **Crop Sci.**, v.37, p.1370-1373, 1997.

VISSCHER, P.M., HALEY, C.S., THOMPSON, R. Marker assisted introgression in backcross breeding programs. **Genetics**, v.144, p.1923-1932, 1996.

ANEXO

Tabela 1A – Simulação do número ótimo de *primers* de SSR por *bootstrap* a partir de 7 *primers* em um total de 57 previamente selecionados, com número de amostras igual a 10, mantendo-se constante o número de genótipos e com um incremento de 1 *primer* a cada simulação.

Nº prim.	Amostra	Correlação	SQDesvio	Estresse
7	1	0,5499	10,8213	0,3196
7	2	0,5947	8,1114	0,2767
7	3	0,4974	10,8868	0,3206
7	4	0,4399	10,7834	0,3191
7	5	0,7015	10,8143	0,3195
7	6	0,4703	9,2412	0,2954
7	7	0,4071	9,6423	0,3017
7	8	0,482	11,6505	0,3316
7	9	0,6518	8,4837	0,283
7	10	0,4787	8,8219	0,2886
7	Média	0,5273	9,9257	0,3056
8	1	0,673	9,0337	0,292
8	2	0,4584	8,8988	0,2898
8	3	0,6979	8,2363	0,2788
8	4	0,669	9,0049	0,2916
8	5	0,5787	7,4512	0,2652
8	6	0,6003	13,2524	0,3537
8	7	0,6613	14,3593	0,3682
8	8	0,6403	7,9886	0,2746
8	9	0,5682	5,7162	0,2323
8	10	0,5301	9,328	0,2968
8	Média	0,6077	9,3269	0,2943
9	1	0,5624	8,1726	0,2778
9	2	0,6956	6,35	0,2448
9	3	0,5584	12,0756	0,3376
9	4	0,5552	7,1282	0,2594
9	5	0,7216	5,103	0,2195
9	6	0,468	8,7642	0,2876
9	7	0,6785	6,7195	0,2519
9	8	0,4718	9,0881	0,2929
9	9	0,7318	12,8636	0,3485
9	10	0,5834	9,2832	0,296
9	Média	0,6027	8,5548	0,2816
10	1	0,5567	5,8821	0,2357
10	2	0,6257	9,9388	0,3063
10	3	0,6119	6,9737	0,2566
10	4	0,6528	5,2211	0,222
10	5	0,6467	5,47	0,2272
10	6	0,6118	5,8334	0,2347
10	7	0,6325	6,9735	0,2566

N° prim.	Amostra	Correlação	SQDesvio	Estresse
10	8	0,5442	6,6527	0,2506
10	9	0,6326	7,5	0,2661
10	10	0,6427	6,9001	0,2552
10	Média	0,6158	6,7345	0,2511
11	1	0,6979	5,064	0,2187
11	2	0,7147	7,066	0,2583
11	3	0,6159	9,1575	0,294
11	4	0,7071	7,9121	0,2733
11	5	0,6715	5,1876	0,2213
11	6	0,6489	4,9193	0,2155
11	7	0,7427	6,0421	0,2388
11	8	0,559	7,0989	0,2589
11	9	0,6901	5,3356	0,2244
11	10	0,6913	6,0218	0,2384
11	Média	0,6739	6,3805	0,2442
12	1	0,7339	6,0525	0,239
12	2	0,7262	4,9926	0,2171
12	3	0,6682	6,8777	0,2548
12	4	0,6805	6,6073	0,2498
12	5	0,7332	5,255	0,2227
12	6	0,7947	8,2701	0,2794
12	7	0,6851	5,7549	0,2331
12	8	0,6946	4,59	0,2082
12	9	0,5954	5,2541	0,2227
12	10	0,7263	6,0538	0,2391
12	Média	0,7038	5,9708	0,2366
13	1	0,6715	4,4378	0,2047
13	2	0,6762	4,964	0,2165
13	3	0,7267	4,4216	0,2043
13	4	0,7201	6,5453	0,2486
13	5	0,7581	3,9821	0,1939
13	6	0,7057	4,6792	0,2102
13	7	0,7032	4,3017	0,2015
13	8	0,6708	7,2962	0,2625
13	9	0,7097	4,3644	0,203
13	10	0,7385	5,0958	0,2193
13	Média	0,708	5,0088	0,2164
14	1	0,7401	5,8725	0,2355
14	2	0,7936	2,8335	0,1636
14	3	0,6988	3,7055	0,187
14	4	0,6922	5,9707	0,2374
14	5	0,6566	4,8799	0,2146
14	6	0,7858	5,9479	0,237
14	7	0,7565	4,8016	0,2129
14	8	0,6471	4,126	0,1974
14	9	0,7723	3,114	0,1715
14	10	0,7711	5,2157	0,2219
14	Média	0,7314	4,6467	0,2079

N° prim.	Amostra	Correlação	SQDesvio	Estresse
15	1	0,7165	4,6527	0,2096
15	2	0,7899	3,3717	0,1784
15	3	0,7242	3,765	0,1885
15	4	0,656	3,7248	0,1875
15	5	0,7977	4,8414	0,2138
15	6	0,7814	3,4409	0,1802
15	7	0,7504	6,2285	0,2425
15	8	0,7519	5,0294	0,2179
15	9	0,7536	3,9055	0,192
15	10	0,8107	2,6779	0,159
15	Média	0,7532	4,1638	0,1969
16	1	0,8133	6,2317	0,2426
16	2	0,7714	5,3898	0,2256
16	3	0,7001	4,0312	0,1951
16	4	0,7826	2,7729	0,1618
16	5	0,6919	7,1828	0,2604
16	6	0,7811	3,4982	0,1817
16	7	0,8085	3,104	0,1712
16	8	0,7166	3,5495	0,1831
16	9	0,7123	3,8323	0,1902
16	10	0,7698	3,0234	0,1689
16	Média	0,7547	4,2616	0,1981
17	1	0,8013	3,4951	0,1816
17	2	0,784	3,2002	0,1738
17	3	0,7838	3,1698	0,173
17	4	0,7174	3,6844	0,1865
17	5	0,7577	4,0778	0,1962
17	6	0,7326	3,131	0,1719
17	7	0,7427	3,3421	0,1776
17	8	0,8001	3,3682	0,1783
17	9	0,7315	3,5786	0,1838
17	10	0,808	4,8435	0,2138
17	Média	0,7659	3,5891	0,1837
18	1	0,7641	5,2093	0,2218
18	2	0,7847	3,8928	0,1917
18	3	0,7427	4,1432	0,1978
18	4	0,8243	2,5758	0,1559
18	5	0,853	4,7815	0,2125
18	6	0,6834	3,0465	0,1696
18	7	0,8097	3,1903	0,1735
18	8	0,8204	2,856	0,1642
18	9	0,7045	3,1697	0,173
18	10	0,765	3,511	0,1821
18	Média	0,7752	3,6376	0,1842
19	1	0,8001	2,6739	0,1589
19	2	0,8204	3,2791	0,1759
19	3	0,8	2,3557	0,1491
19	4	0,8624	3,3945	0,179

N° prim.	Amostra	Correlação	SQDesvio	Estresse
19	6	0,7845	3,6938	0,1867
19	7	0,7497	2,7873	0,1622
19	8	0,8693	1,9533	0,1358
19	9	0,7671	2,7834	0,1621
19	10	0,7806	3,0855	0,1707
19	Média	0,8038	2,9544	0,1663
20	1	0,8399	2,2792	0,1467
20	2	0,8215	2,2009	0,1441
20	3	0,8126	3,1253	0,1718
20	4	0,8662	1,848	0,1321
20	5	0,7926	2,7797	0,162
20	6	0,7444	2,66	0,1585
20	7	0,775	3,5547	0,1832
20	8	0,7808	2,9748	0,1676
20	9	0,7753	2,7921	0,1624
20	10	0,7857	2,4206	0,1512
20	Média	0,7994	2,6635	0,1579
21	1	0,7799	2,7034	0,1598
21	2	0,7833	2,9605	0,1672
21	3	0,7984	2,9655	0,1673
21	4	0,8714	2,0404	0,1388
21	5	0,8334	2,1113	0,1412
21	6	0,8121	2,4121	0,1509
21	7	0,7639	3,0165	0,1688
21	8	0,8365	2,8427	0,1638
21	9	0,7941	2,566	0,1556
21	10	0,8088	2,8287	0,1634
21	Média	0,8082	2,6447	0,1577
22	1	0,7924	2,7744	0,1618
22	2	0,8177	2,1645	0,143
22	3	0,819	2,2481	0,1457
22	4	0,8638	2,1629	0,1429
22	5	0,8326	2,0166	0,138
22	6	0,7713	2,7614	0,1615
22	7	0,8081	2,552	0,1552
22	8	0,821	1,9612	0,1361
22	9	0,8449	1,9728	0,1365
22	10	0,8389	2,4772	0,1529
22	Média	0,821	2,3091	0,1474
23	1	0,8526	2,4551	0,1522
23	2	0,8675	1,866	0,1327
23	3	0,8306	2,3714	0,1496
23	4	0,8876	2,364	0,1494
23	5	0,8823	1,583	0,1222
23	6	0,834	2,0177	0,138
23	7	0,8542	3,0936	0,1709
23	8	0,8232	3,1653	0,1729
23	9	0,845	2,4879	0,1533

N° prim.	Amostra	Correlação	SQDesvio	Estresse
23	10	0,7238	3,0721	0,1703
23	Média	0,8401	2,4476	0,1512
24	1	0,8083	2,0662	0,1397
24	2	0,8755	2,0374	0,1387
24	3	0,8258	2,731	0,1606
24	4	0,8475	2,3616	0,1493
24	5	0,7912	2,1502	0,1425
24	6	0,887	2,8406	0,1638
24	7	0,8506	1,9299	0,135
24	8	0,8576	2,4687	0,1527
24	9	0,8579	1,561	0,1214
24	10	0,8262	2,454	0,1522
24	Média	0,8428	2,2601	0,1456
25	1	0,8586	1,7989	0,1303
25	2	0,8847	1,8655	0,1327
25	3	0,8469	1,8527	0,1323
25	4	0,8337	1,7005	0,1267
25	5	0,8673	1,7612	0,1289
25	6	0,8696	1,6622	0,1253
25	7	0,8317	2,2813	0,1468
25	8	0,8315	1,9659	0,1362
25	9	0,8186	2,1088	0,1411
25	10	0,7778	2,3608	0,1493
25	Média	0,842	1,9358	0,135
26	1	0,8429	2,0765	0,14
26	2	0,8264	2,0494	0,1391
26	3	0,8672	1,3455	0,1127
26	4	0,7761	2,0487	0,1391
26	5	0,8907	1,4354	0,1164
26	6	0,8411	3,2734	0,1758
26	7	0,8809	1,503	0,1191
26	8	0,8442	2,5633	0,1556
26	9	0,8361	1,7847	0,1298
26	10	0,8548	1,8772	0,1331
26	Média	0,846	1,9957	0,1361
27	1	0,8824	1,6755	0,1258
27	2	0,8822	1,5625	0,1215
27	3	0,8849	1,7138	0,1272
27	4	0,9169	1,4795	0,1182
27	5	0,8776	2,1586	0,1428
27	6	0,878	1,4212	0,1158
27	7	0,858	2,1321	0,1419
27	8	0,8908	1,4225	0,1159
27	9	0,8622	2,1868	0,1437
27	10	0,8517	1,5829	0,1222
27	Média	0,8785	1,7335	0,1275
28	1	0,858	1,7648	0,1291
28	2	0,8671	2,4386	0,1517

N° prim.	Amostra	Correlação	SQDesvio	Estresse
28	3	0,8593	2,717	0,1602
28	4	0,873	1,363	0,1134
28	5	0,8625	2,8953	0,1653
28	6	0,8536	2,8984	0,1654
28	7	0,899	1,5056	0,1192
28	8	0,9004	1,8896	0,1336
28	9	0,8782	1,7156	0,1273
28	10	0,8898	1,3475	0,1128
28	Média	0,8741	2,0535	0,1378
29	1	0,8853	1,6823	0,126
29	2	0,8762	1,5873	0,1224
29	3	0,8412	1,7953	0,1302
29	4	0,8651	1,4245	0,116
29	5	0,8661	1,4251	0,116
29	6	0,9113	1,0202	0,0981
29	7	0,8605	1,4676	0,1177
29	8	0,88	1,4585	0,1173
29	9	0,8705	1,6504	0,1248
29	10	0,8883	1,407	0,1153
29	Média	0,8745	1,4918	0,1184
30	1	0,8838	1,2816	0,11
30	2	0,8625	1,5396	0,1206
30	3	0,8347	1,669	0,1255
30	4	0,9001	1,1138	0,1025
30	5	0,8943	1,2952	0,1106
30	6	0,8789	1,4417	0,1167
30	7	0,9069	1,8116	0,1308
30	8	0,8679	1,6457	0,1246
30	9	0,889	1,2511	0,1087
30	10	0,902	1,3873	0,1144
30	Média	0,882	1,4437	0,1164
31	1	0,9171	1,0772	0,1008
31	2	0,9136	1,163	0,1048
31	3	0,864	1,4289	0,1161
31	4	0,8941	1,4107	0,1154
31	5	0,8912	1,4344	0,1164
31	6	0,8675	1,8568	0,1324
31	7	0,8818	1,317	0,1115
31	8	0,8669	1,4852	0,1184
31	9	0,8612	1,3359	0,1123
31	10	0,9183	0,9134	0,0929
31	Média	0,8876	1,3423	0,1121
32	1	0,9245	1,097	0,1018
32	2	0,9014	1,0278	0,0985
32	3	0,9113	1,5233	0,1199
32	4	0,8907	1,1717	0,1052
32	5	0,9107	1,0554	0,0998
32	6	0,8891	1,1821	0,1056

N° prim.	Amostra	Correlação	SQDesvio	Estresse
32	7	0,9119	1,3637	0,1135
32	8	0,8859	1,5396	0,1206
32	9	0,8804	1,1559	0,1045
32	10	0,8847	1,0921	0,1015
32	Média	0,8991	1,2209	0,1071
33	1	0,9215	0,8918	0,0918
33	2	0,9101	1,2778	0,1098
33	3	0,9067	1,0042	0,0974
33	4	0,903	1,0514	0,0996
33	5	0,9035	1,1204	0,1028
33	6	0,9076	1,0244	0,0983
33	7	0,9311	0,7584	0,0846
33	8	0,8983	0,9697	0,0957
33	9	0,9237	0,7912	0,0864
33	10	0,9193	1,0854	0,1012
33	Média	0,9125	0,9975	0,0968
34	1	0,923	1,2715	0,1096
34	2	0,9204	1,0608	0,1001
34	3	0,9298	0,7509	0,0842
34	4	0,9144	1,7061	0,1269
34	5	0,8972	0,9648	0,0954
34	6	0,8978	1,8202	0,1311
34	7	0,9286	0,8997	0,0922
34	8	0,9012	1,4104	0,1154
34	9	0,9199	0,8064	0,0873
34	10	0,9154	1,9582	0,136
34	Média	0,9148	1,2649	0,1078
35	1	0,9029	0,9148	0,0929
35	2	0,9495	0,518	0,0699
35	3	0,9062	1,1241	0,103
35	4	0,9285	1,1733	0,1052
35	5	0,9335	1,0663	0,1003
35	6	0,8751	1,4408	0,1166
35	7	0,9225	0,9765	0,096
35	8	0,9314	0,9662	0,0955
35	9	0,9389	0,937	0,0941
35	10	0,9229	1,0654	0,1003
35	Média	0,9212	1,0182	0,0974
36	1	0,9264	0,9302	0,0937
36	2	0,8978	0,9869	0,0965
36	3	0,9508	1,6642	0,1253
36	4	0,893	1,2518	0,1087
36	5	0,9295	0,8455	0,0893
36	6	0,9332	0,7925	0,0865
36	7	0,9389	0,6358	0,0775
36	8	0,9254	0,7953	0,0866
36	9	0,901	1,0924	0,1016
36	10	0,9301	0,7364	0,0834

N° prim.	Amostra	Correlação	SQDesvio	Estresse
36	Média	0,9226	0,9731	0,0949
37	1	0,9408	0,5864	0,0744
37	2	0,9223	0,7681	0,0852
37	3	0,9198	0,8371	0,0889
37	4	0,9419	0,6039	0,0755
37	5	0,9272	0,8042	0,0871
37	6	0,9162	0,8842	0,0914
37	7	0,9377	0,8474	0,0894
37	8	0,9376	0,7222	0,0826
37	9	0,9279	0,8131	0,0876
37	10	0,9413	0,6331	0,0773
37	Média	0,9313	0,75	0,0839
38	1	0,9281	0,8378	0,0889
38	2	0,9237	0,8269	0,0884
38	3	0,9412	0,7314	0,0831
38	4	0,9336	0,8321	0,0886
38	5	0,9221	0,8645	0,0903
38	6	0,9356	0,6462	0,0781
38	7	0,9214	0,8262	0,0883
38	8	0,9389	0,6479	0,0782
38	9	0,9336	0,6877	0,0806
38	10	0,9218	0,8567	0,0899
38	Média	0,93	0,7757	0,0855
39	1	0,9332	0,8088	0,0874
39	2	0,9398	0,5869	0,0744
39	3	0,9398	0,5915	0,0747
39	4	0,9317	0,8112	0,0875
39	5	0,9355	1,2059	0,1067
39	6	0,943	0,7032	0,0815
39	7	0,9438	0,7381	0,0835
39	8	0,94	0,6827	0,0803
39	9	0,937	0,6613	0,079
39	10	0,9299	0,6826	0,0803
39	Média	0,9374	0,7472	0,0835
40	1	0,9546	0,472	0,0668
40	2	0,9525	0,4909	0,0681
40	3	0,944	0,675	0,0798
40	4	0,9449	0,6398	0,0777
40	5	0,955	0,936	0,094
40	6	0,9481	0,6525	0,0785
40	7	0,936	0,726	0,0828
40	8	0,9545	0,6236	0,0767
40	9	0,9484	0,5445	0,0717
40	10	0,9539	0,5285	0,0706
40	Média	0,9492	0,6289	0,0767
41	1	0,9521	0,705	0,0816
41	2	0,9481	0,512	0,0695
41	3	0,9522	0,4948	0,0683

N° prim.	Amostra	Correlação	SQDesvio	Estresse
41	4	0,9541	0,4975	0,0685
41	5	0,9603	0,4355	0,0641
41	6	0,952	0,5301	0,0707
41	7	0,9553	0,4798	0,0673
41	8	0,9545	0,4906	0,0681
41	9	0,9451	0,5466	0,0718
41	10	0,9465	0,6069	0,0757
41	Média	0,952	0,5299	0,0706
42	1	0,9514	0,5449	0,0717
42	2	0,9499	0,5032	0,0689
42	3	0,9582	0,4067	0,062
42	4	0,9615	0,3971	0,0612
42	5	0,9561	0,5127	0,0696
42	6	0,9512	0,9936	0,0969
42	7	0,9427	0,5724	0,0735
42	8	0,9399	0,8994	0,0921
42	9	0,9485	0,5094	0,0694
42	10	0,9536	0,4906	0,0681
42	Média	0,9513	0,583	0,0733
43	1	0,9528	0,46	0,0659
43	2	0,9396	0,6644	0,0792
43	3	0,9497	0,481	0,0674
43	4	0,9527	0,598	0,0751
43	5	0,9567	0,5347	0,071
43	6	0,955	0,5089	0,0693
43	7	0,9599	0,4407	0,0645
43	8	0,9539	0,455	0,0655
43	9	0,9611	0,3712	0,0592
43	10	0,9545	0,5182	0,0699
43	Média	0,9536	0,5032	0,0687
44	1	0,954	0,54	0,0714
44	2	0,9548	0,6554	0,0787
44	3	0,9694	0,3623	0,0585
44	4	0,9695	0,3746	0,0595
44	5	0,967	0,5341	0,071
44	6	0,9631	0,4482	0,0651
44	7	0,9696	0,398	0,0613
44	8	0,9661	0,3428	0,0569
44	9	0,9591	0,4553	0,0656
44	10	0,9671	0,3394	0,0566
44	Média	0,964	0,445	0,0644
45	1	0,958	0,6032	0,0755
45	2	0,9688	0,3672	0,0589
45	3	0,9676	0,3278	0,0556
45	4	0,9702	0,324	0,0553
45	5	0,9679	0,3694	0,0591
45	6	0,97	0,2924	0,0525
45	7	0,963	0,4168	0,0627

N° prim.	Amostra	Correlação	SQDesvio	Estresse
45	8	0,9728	0,3174	0,0547
45	9	0,9656	0,3641	0,0586
45	10	0,9726	0,2745	0,0509
45	Média	0,9677	0,3657	0,0584
46	1	0,9525	0,467	0,0664
46	2	0,9701	0,2993	0,0532
46	3	0,9769	0,3068	0,0538
46	4	0,9715	0,3707	0,0592
46	5	0,9642	0,3917	0,0608
46	6	0,9605	0,3818	0,06
46	7	0,9693	0,3207	0,055
46	8	0,9555	0,4229	0,0632
46	9	0,973	0,268	0,0503
46	10	0,9726	0,2777	0,0512
46	Média	0,9666	0,3507	0,0573
47	1	0,976	0,4184	0,0628
47	2	0,9666	0,3298	0,0558
47	3	0,9692	0,3157	0,0546
47	4	0,9693	0,3495	0,0574
47	5	0,9701	0,2933	0,0526
47	6	0,9746	0,2689	0,0504
47	7	0,9784	0,3018	0,0534
47	8	0,9707	0,3302	0,0558
47	9	0,9704	0,2996	0,0532
47	10	0,962	0,3863	0,0604
47	Média	0,9707	0,3294	0,0556
48	1	0,9834	0,1856	0,0419
48	2	0,9737	0,2889	0,0522
48	3	0,976	0,2646	0,05
48	4	0,9742	0,3535	0,0578
48	5	0,9736	0,2533	0,0489
48	6	0,9787	0,2111	0,0446
48	7	0,9684	0,3218	0,0551
48	8	0,9748	0,2538	0,0489
48	9	0,9733	0,2688	0,0504
48	10	0,9727	0,3332	0,0561
48	Média	0,9749	0,2734	0,0506
49	1	0,9758	0,2591	0,0495
49	2	0,9786	0,2919	0,0525
49	3	0,9762	0,2519	0,0488
49	4	0,9772	0,2238	0,046
49	5	0,9752	0,2747	0,0509
49	6	0,9769	0,3162	0,0546
49	7	0,9679	0,3093	0,054
49	8	0,9789	0,2558	0,0491
49	9	0,978	0,2211	0,0457
49	10	0,9741	0,3162	0,0546
49	Média	0,9759	0,272	0,0506

N° prim.	Amostra	Correlação	SQDesvio	Estresse
50	1	0,9775	0,2252	0,0461
50	2	0,9781	0,2095	0,0445
50	3	0,9828	0,1833	0,0416
50	4	0,9851	0,1474	0,0373
50	5	0,9795	0,2583	0,0494
50	6	0,9823	0,2737	0,0508
50	7	0,9792	0,2148	0,045
50	8	0,9842	0,1632	0,0393
50	9	0,9804	0,1904	0,0424
50	10	0,9836	0,1734	0,0405
50	Média	0,9813	0,2039	0,0437
51	1	0,9866	0,2317	0,0468
51	2	0,9842	0,1753	0,0407
51	3	0,9757	0,2931	0,0526
51	4	0,9849	0,166	0,0396
51	5	0,9839	0,1632	0,0392
51	6	0,9839	0,1724	0,0403
51	7	0,9861	0,1441	0,0369
51	8	0,9845	0,1713	0,0402
51	9	0,9891	0,2625	0,0498
51	10	0,9884	0,2363	0,0472
51	Média	0,9847	0,2016	0,0433
52	1	0,9914	0,0854	0,0284
52	2	0,9868	0,1652	0,0395
52	3	0,9868	0,1411	0,0365
52	4	0,9852	0,1575	0,0386
52	5	0,9877	0,133	0,0354
52	6	0,9885	0,1214	0,0338
52	7	0,9882	0,1752	0,0407
52	8	0,9858	0,1561	0,0384
52	9	0,9845	0,1811	0,0413
52	10	0,9884	0,1412	0,0365
52	Média	0,9873	0,1457	0,0369
53	1	0,9896	0,1137	0,0328
53	2	0,9893	0,1037	0,0313
53	3	0,9879	0,1231	0,0341
53	4	0,9889	0,1132	0,0327
53	5	0,9915	0,0869	0,0286
53	6	0,989	0,1254	0,0344
53	7	0,9915	0,0952	0,03
53	8	0,9882	0,1522	0,0379
53	9	0,9902	0,104	0,0313
53	10	0,9877	0,123	0,0341
53	Média	0,9894	0,114	0,0327
54	1	0,9929	0,0746	0,0265
54	2	0,993	0,0825	0,0279
54	3	0,993	0,0907	0,0293
54	4	0,991	0,0934	0,0297

N° prim.	Amostra	Correlação	SQDesvio	Estresse
54	5	0,9934	0,0795	0,0274
54	6	0,9927	0,0925	0,0295
54	7	0,9917	0,0939	0,0298
54	8	0,9924	0,0745	0,0265
54	9	0,9914	0,091	0,0293
54	10	0,9914	0,0893	0,029
54	Média	0,9923	0,0862	0,0285
55	1	0,9929	0,0749	0,0266
55	2	0,9946	0,0557	0,0229
55	3	0,9948	0,0569	0,0232
55	4	0,9952	0,0549	0,0228
55	5	0,9945	0,0544	0,0227
55	6	0,9956	0,0457	0,0208
55	7	0,9942	0,0637	0,0245
55	8	0,9952	0,0483	0,0214
55	9	0,9931	0,0667	0,0251
55	10	0,9921	0,0782	0,0272
55	Média	0,9942	0,0599	0,0237
56	1	0,9974	0,0261	0,0157
56	2	0,9983	0,019	0,0134
56	3	0,9969	0,0434	0,0202
56	4	0,9973	0,033	0,0176
56	5	0,9971	0,0296	0,0167
56	6	0,9967	0,0316	0,0173
56	7	0,9974	0,0318	0,0173
56	8	0,9967	0,0316	0,0173
56	9	0,9961	0,0388	0,0191
56	10	0,9966	0,0348	0,0181
56	Média	0,997	0,032	0,0173
57		1,0	0,0	0,0

Tabela 2A – Simulação dos agrupamentos dos 27 genótipos com o número ótimo de 44 *primers* - amostra número 1. *Primers* amostrados foram: 4-5-6-7-8-9-10-12-13-14-15-16-18-20-21-25-27-29-30-31-32-33-34-35-36-37-39-40-41-42-43-45-46-47-48-49-50-51-53-54-55-56-57.

Grupo	Genótipos
1	3 4 5 6 8 9 7
2	11 12 1 27 15 25
3	18 21 22
4	13 16 17
5	10 14 20 2
6	26
7	19
8	24
9	23

Tabela 3A – Simulação dos agrupamentos dos 27 genótipos com o número ótimo de 44 *primers* - amostra número 2. *Primers* amostrados foram: 1-2-3-4-5-6-7-10-12-13-15-16-17-18-19-20-21-22-25-26-27-29-30-32-33-34-35-37-38-39-40-41-42-43-44-45-46-47-48-50-52-53-54-55.

Grupo	Genótipos
1	3 4 5 6 8 9 7
2	11 12 15 27 2 10
3	18 21 22
4	16 17 13
5	1 25
6	14 20
7	19 24
8	23
9	26

Tabela 4A – Simulação dos agrupamentos dos 27 genótipos com o número ótimo de 44 *primers* - amostra número 3. *Primers* amostrados foram: 4-5-6-7-8-11-12-13-15-16-17-18-20-21-22-23-24-25-26-27-28-30-32-34-35-36-37-38-39-40-42-43-44-45-46-48-49-50-51-52-53-54-55-56.

Grupo	Genótipos
1	3 4 5 6 8 7 9
2	11 12 10 27 2 15 19
3	18 21 22
4	16 17 13
5	24 25 1
6	20 26
7	14
8	23

Tabela 5A – Simulação dos agrupamentos dos 27 genótipos com o número ótimo de 44 *primers* - amostra número 4. *Primers* amostrados foram: 1-4-5-6-7-8-9-10-11-13-14-17-18-20-21-22-23-24-26-27-28-29-30-31-32-33-34-36-37-38-39-40-41-42-43-44-46-47-48-50-51-52-55-56.

Grupo	Genótipos
1	3 4 5 8 6 9 7
2	11 12 27 15 10 14 2
3	18 21 22
4	16 17 13
5	19 24
6	1 25
7	23
8	26
9	20

Tabela 6A – Simulação dos agrupamentos dos 27 genótipos com o número ótimo de 44 *primers* - amostra número 5. *Primers* amostrados foram: 1-2-3-4-5-6-7-9-10-11-12-13-14-15-16-18-20-21-22-23-24-25-27-28-29-30-32-34-35-36-37-39-40-42-44-47-48-49-50-51-54-56-57.

Grupo	Genótipos
1	3 4 5 6 8 9 7
2	11 12 10 27 14 2 15
3	18 21 22
4	13 16 17
5	1 25
6	19 24
7	23
8	20
9	26

Tabela 7A – Simulação dos agrupamentos dos 27 genótipos com o número ótimo de 44 *primers* - amostra número 6. *Primers* amostrados foram: 1-2-3-4-5-6-8-11-12-13-14-15-17-18-20-21-22-23-25-26-28-29-31-32-33-34-35-36-37-38-40-41-42-43-44-45-46-49-50-51-53-54-55-57.

Grupo	Genótipos
1	11 12 15 27 10 14 2 1
2	3 4 5 6 8 9 7
3	18 21 22
4	13 16 17
5	25 26
6	19 24
7	23
8	20

Tabela 8A – Simulação dos agrupamentos dos 27 genótipos com o número ótimo de 44 *primers* - amostra número 7. *Primers* amostrados foram: 1-2-3-4-5-8-10-11-12-13-14-17-20-21-22-23-24-25-26-27-28-29-30-31-32-33-34-35-36-37-39-40-41-42-43-45-47-48-51-52-54-55-56-57.

Grupo	Genótipos
1	3 4 5 6 8 9 7
2	11 12 27 15 1 25
3	18 21 22
4	10 14 13 20
5	16 17
6	19 24
7	23
8	26
9	2

Tabela 9A – Simulação dos agrupamentos dos 27 genótipos com o número ótimo de 44 *primers* - amostra número 8. *Primers* amostrados foram: 1-2-4-5-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-23-24-25-26-28-29-30-31-32-34-36-37-39-40-43-44-46-49-50-51-53-54-55-56-57.

Grupo	Genótipos
1	3 4 5 6 8 9 7
2	11 12 15 27 10 14
3	18 21 22
4	13 16 17
5	23 24 19
6	1 25
7	26
8	20
9	2

Tabela 10A – Simulação dos agrupamentos dos 27 genótipos com o número ótimo de 44 *primers* - amostra número 9. *Primers* amostrados foram: 1-3-4-5-6-10-11-12-16-20-21-22-23-24-25-26-27-28-29-30-31-32-33-34-36-37-38-39-40-41-43-44-45-46-47-48-49-50-51-52-54-55-56-57.

Grupo	Genótipos
1	3 4 5 8 6 9 7
2	11 12 10 2 27 14 15
3	18 21 22
4	16 17 13
5	1 25
6	19 24
7	23
8	20
9	26

Tabela 11A – Simulação dos agrupamentos dos 27 genótipos com o número ótimo de 44 *primers* - amostra número 10. *Primers* amostrados foram: 1-2-3-4-8-9-10-12-13-15-16-17-19-20-21-24-25-26-27-29-30-31-33-34-35-36-37-39-42-43-44-45-46-47-48-49-50-51-52-53-54-55-56-57.

Grupo	Genótipos
1	3 4 5 8 6 9 7
2	11 12 15 27 2
3	18 21 22
4	10 14 13
5	16 17
6	1 25 24
7	19 23
8	26
9	20

CAPÍTULO 3

INTROGRESSÃO DE ALELOS PARA ALTO TEOR DE PROTEÍNA EM VARIEDADES DE SOJA

RESUMO

No Brasil, os programas de melhoramento de soja, até então, priorizavam o aumento de produtividade e do teor de óleo. Estes dois caracteres apresentam correlação negativa com teor de proteína. Ao longo de sucessivos ciclos de seleção o teor de proteína foi sendo reduzido nas variedades comerciais. Atualmente, o farelo de soja é um dos principais produtos das indústrias de esmagamento de soja no Brasil. Uma vez que o teor de proteína no farelo influencia a formação do preço, torna-se necessário o desenvolvimento de variedades com alto teor de proteína e produtivas. No entanto, as fontes de alelos para alto teor de proteína são introduções de plantas ou germoplasma exótico o que seria um retrocesso em programas de melhoramento em termos de ganhos genéticos para produtividade. A correlação negativa geralmente encontrada entre teor de proteína e produtividade das populações melhoradas é atribuída ao fato do parental de alto teor de proteína sempre apresentar baixa produtividade e ao número insuficiente de ciclos de seleção que são realizados. Uma alternativa para reduzir esta associação negativa entre teor de proteína e produtividade é o desenvolvimento de germoplasmas adaptados, com alto teor de proteína e produtividade similar ao progenitor recorrente, por meio de retrocruzamentos, os quais poderão ser utilizados a seguir em programas de melhoramento que visem a introdução de alto teor de proteína em variedades comerciais. Neste sentido, os objetivos deste trabalho foram: (a) identificar cruzamentos com melhor capacidade de combinação para alto teor de proteína; (b) identificar os melhores progenitores doadores; (c) efetuar seleção individual baseada em microanálises bioquímicas não destrutivas na geração F₂ dos melhores cruzamentos; (d) avaliar os ganhos de seleção em F₃, com seleção em F₂; (e) efetuar *fingerprinting* dos indivíduos selecionados para identificar os mais próximos do progenitor recorrente e (f) avaliar a eficiência de seleção com a utilização de marcadores moleculares. Sete variedades comerciais oriundas da COODETEC foram cruzadas com dezoito linhagens de alto teor de proteína do programa de melhoramento da qualidade da soja para a agroindústria do BIOAGRO/UFV. O delineamento genético utilizado foi o modelo hierárquico

onde os pais eram as variedades elite (recorrentes) e as mães eram as linhagens de alto teor de proteína (doadores). A análise de variância evidenciou variabilidade genética em todos os cruzamentos e um alto valor de herdabilidade para teor de proteína (88,7%). O progenitor recorrente OC953312 apresentou maior efeito de capacidade geral de combinação (CGC), e em seus cruzamentos detectou-se maior variabilidade genética (4,62) e maior herdabilidade (93,1%) para teor de proteína. Os melhores progenitores doadores foram escolhidos com base no teste de média a 5% de probabilidade, sendo eles: OC128PT258, CD203PT30-3, CD203PT30-2, CD203PT31-3, CD202PT31-3, CD202PT31-2 e CD201PT82-3. A seleção desses progenitores doadores como sendo os melhores para dar prosseguimento a outros programas de melhoramento para alto teor de proteína, diminuirá a necessidade do programa de melhoramento de soja do BIOAGRO/UFV de utilizar um grande número de progenitores doadores, reduzindo, conseqüentemente, o número de cruzamentos, o número de análises e os custos gerais do programa. O teor de proteína de plantas F3 derivadas de sementes F2 selecionadas foi confirmado pelo método de Kjeldahl, apresentando em média 43,03%. Os ganhos, para alto teor de proteína, absoluto e percentual das plantas selecionadas, foram de 5,68, e 15,2%, respectivamente e a distância genética em relação ao progenitor recorrente CD 98 5015 foi em média de 0,44. As plantas selecionadas derivadas dos progenitores recorrentes CD 204, CD210, CD211, OC953006, OC953312 e CD983343 tiveram ganhos médios reais de: 8,41 3,88, 9,35 6,58, 6,75, 8,87, respectivamente. As plantas selecionadas para alto teor de proteína serão analisadas quanto às suas distâncias genéticas em relação aos seus progenitores recorrentes por meio de marcadores microssatélites. As plantas com maior teor de proteína e menor distância em relação ao progenitor recorrente serão selecionadas para dar prosseguimento ao programa de introgressão de alelos de alto teor de proteína por meio de retrocruzamentos assistido por marcadores moleculares.

Palavras chave: Retrocruzamento, seleção assistida, marcadores moleculares, melhoramento, correlação.

INTRODUÇÃO

O farelo de soja destinado à exportação é classificado, de acordo com o seu conteúdo de proteína, nas categorias “HyPro” (> 48%), “Normal” (46%) e “LowPro” (< 43,5%). Lotes classificados como HyPro recebem ágio sobre o valor normal de comercialização, enquanto que lotes classificados com “LowPro” sofrem deságio sobre este valor. Para atingir o nível classificado como “Normal” e “HyPro” a soja deve conter, no mínimo, 41,5% e 43% de proteína nas sementes, respectivamente, com base na matéria seca (informações obtidas junto a Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda – COODETEC, Cascavel, PR). A grande maioria das variedades brasileiras de soja, em especial aquelas cultivadas na região sul, possui teor de proteína considerado baixo, isto porque, os programas de melhoramento, ao longo dos anos, selecionaram linhagens apenas em função da característica produtividade, a qual possui correlação negativa com o conteúdo de proteína dos grãos. O desenvolvimento de cultivares com alto teor de proteína e alta produtividade é mais trabalhoso devido à correlação negativa entre teor de proteína e produtividade. WILCOX e CAVINS (1995) encontraram em uma revisão sobre trabalhos de melhoramento uma correlação negativa entre proteína e produtividade de moderada a severa, a qual variou de – 0,23 a – 0,86.

Durante muito tempo o número de genótipos que poderia ser avaliado para teor de proteína era limitado pelo tempo de análise, pelo custo e pela quantidade de material necessário, uma vez que o principal método de quantificação de proteína total era restrito ao método de Kjeldahl, o qual é um método destrutivo de análise. A adaptação de uma metodologia de quantificação de proteína que possibilitasse a seleção de sementes após a primeira geração de autofecundação sem prejudicar o poder germinativo da semente seria extremamente útil, como é o exemplo do método do ácido bicinconínico (SMITH 1985). Vários autores têm sugerido diversas alternativas para seleção para alto teor de proteína. Neste sentido, MILLER e FEHR (1979) avaliaram a eficiência do melhoramento por meio de seleção recorrente praticando seleção direta para alto teor de proteína e indireta baseada no baixo teor de óleo. Também FEHR e WEBER (1968) sugeriram

a seleção para alto teor de proteína com base na seleção de sementes de tamanho grande e alta densidade específica. ISHIGE (1984) utilizou o método modificado do biureto para analisar sementes individuais e estimar parâmetros genéticos de populações segregantes para proteína em soja.

ERICKSON *et al.* (1981) avaliaram o uso de seleção para proteína em gerações precoce do cruzamento entre *Glycine soja* e *Glycine max*. A herdabilidade para teor de proteína foi medida em F2 e F3 e a seleção massal e a seleção de famílias foram comparadas na geração F3 destes cruzamentos. *G. soja* foi usado como progenitor masculino em todos os cruzamentos e a porcentagem de seleção foi de 10%. A herdabilidade para proteína na geração F2 foi de 27% obtida pela regressão pai-filho. A herdabilidade no sentido amplo em famílias F3 cultivadas em dois locais, em uma época de cultivo, foi de 78%. A média do teor de proteína em cada população selecionada foi maior do que a média da população não selecionada (controle). As médias dos teores de proteína obtida nos diferentes pelos métodos de seleção foram: seleção massal em F2 e F3 48%; seleção massal em F3 seguido de seleção de uma única semente por descendência 48,8%; seleção entre famílias F3 47,6%; seleção entre e dentro de famílias F3 47,5%. Concluiu-se que a média da população melhorada pelo método de seleção massal foi superior à média obtida pelo método de seleção de famílias em nível de 1% de probabilidade. Estes autores comentam que apesar de a seleção precoce ter elevado o teor de proteína da população, a segregação continuou a originar genótipos de baixo teor de proteína, conseqüentemente, passos adicionais de seleção seriam necessários.

HARTWIG (1994) demonstrou que progressos podem ser obtidos na seleção de linhagens de soja de alto teor de proteína e alta produtividade quando se utiliza o progenitor de alto teor de proteína com tipo agrônomo melhorado e com produtividade média.

Os objetivos deste trabalho foram: (a) identificar cruzamentos com melhor capacidade de combinação para alto teor de proteína; (b) identificar os melhores progenitores doadores de alelos para alto teor de proteína; (c) efetuar seleção individual baseada em microanálises bioquímicas na geração F2 dos melhores cruzamentos; (d) avaliar os ganhos de seleção em

F3 com seleção em F2; (e) efetuar *fingerprinting* dos indivíduos selecionados para identificar os mais próximos do progenitor recorrente e (f) avaliar a eficiência de seleção com a utilização de marcadores moleculares.

MATERIAL E MÉTODOS

Material genético

Três variedades (CD 204, CD 210, CD 211) e seis linhagens de soja (CD 983321, CD 985015, CD 983343, OC 953006, OC 953194, e OC 953312) provenientes da Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda – COODETEC constantes da Tabela 1 (exceto CD 985015 que já estava em um estágio mais avançado de cruzamento) foram utilizadas como progenitores recorrentes. Dezoito linhagens, com alto teor de proteína do programa de melhoramento da soja do BIOAGRO/UFV, foram utilizadas como fontes doadoras da característica alto teor de proteína com o objetivo de se incorporar alelos para alto teor de proteína nas variedades da COODETEC.

Inicialmente, foram efetuados cruzamentos entre os progenitores (recorrentes e doadores), obtendo-se as combinações híbridas (sementes F1). Todas essas combinações híbridas foram autofecundadas em casa de vegetação para obter as progênies F2 (populações segregantes). Foram analisadas 400 sementes segregantes obtidas de cada população. Para escolha dos cruzamentos mais promissores foram selecionados aqueles que apresentaram maior média e maior variância genética para conteúdo de proteína na geração F2.

Como critério para escolha dos melhores progenitores doadores, foram escolhidos o que obtiveram as maiores médias para teor de proteína quando avaliados em conjunto pelo teste de média. Individualmente, cada semente F2, em um total de 400 sementes, foi submetida a uma microanálise não destrutiva para teor de proteína pelo método que utiliza o ácido bicinconínico (BCA) (SMITH *et al.* 1985) adaptado para análises de proteína em sementes de soja no laboratório de genética molecular de plantas – Proteína/BIOAGRO. Com a finalidade de selecionar os genótipos F2 com maiores teores de proteína foi utilizada uma porcentagem de seleção de 10%. Com isso, 40 sementes F2 foram selecionadas. e plantadas, obtendo-se plantas F2. Foi extraído DNA das plantas F2 para análises com marcadores microssatélites (*fingerprinting*). A seguir, foram

escolhidas as plantas geneticamente mais próximas do progenitor recorrente para efetuar o primeiro retrocruzamento, obtendo-se com isso as progênes RC1F1.

Tabela 1 – Cruzamentos obtidos entre os progenitores doadores e os progenitores recorrentes para teor de proteína em soja.

Progenitor recorrente	Progenitor doador	Número de repetições do cruzamento
CD204	OC671PT46-2	6
	CD203PT30-1	7
	CD203PT30-2	9
	OC671PT100-1	6
	CD202PT31-1	3
	CD202PT31-3	5
	OC671PT46-1	5
	CD202PT111-1	3
	OC128PT258	4
	OC13PT158-3	2
CD210	CD202PT31-3	2
	CD202PT31-2	8
	OC671PT46-2	4
	OC671PT100-1	2
	CD201PT82-3	2
CD211	CD203PT30-1	2
	CD203PT30-2	2
	CD202PT111-1	3
	OC671PT46-1	15
	CD203PT30-1	2
CD983343	CD203PT30-2	4
	CD202PT111-1	2
	OC128PT258	6
	OC671PT46-1	6
	CD203PT30-3	6
OC953006	CD202PT31-1	4
	CD202PT31-3	4
	OC671PT46-2	3
	CD202PT31-2	2
	CD203PT30-2	10
	CD202PT111-1	6
	CD201PT82-3	7
	OC671PT46-1	10
	OC14PT260	7
	OC953194	OC671PT46-2
OC671PT46-1		9
CD203PT30-2		6
OC953312	OC671PT46-2	11
	OC13PT158-3	14
	CD203PT30-2	10
	CD202PT31-3	4
	CD201PT82-3	4
	CD203PT30-3	4
	CD202PT111-1	7
	OC671PT46-1	6
	OC128PT258-2	2
	CD203PT30-1	4

Análise do teor de proteína

O teor de proteína das sementes individuais da geração F2 foi determinado pelo método do ácido bicinconínico, conforme descrito por SMITH *et al.* (1985), com modificações. De cada semente F2, eram cortados 10 mg de tecido do cotilédone, no lado oposto ao embrião. A porção cortada das sementes era macerada em gral na presença de 2 mL de uma solução de SDS 1%. Depois de homogeneizada, esta solução era diluída cerca de 300 vezes e alíquotas de 100 µL de cada tubo eram transferidas para placas de ELISA (3 repetições). A seguir, eram adicionados 200 µL de solução de ácido bicinconínico contendo 2% de sulfato de cobre. Em cada placa, eram aplicadas também quantidades crescentes e conhecidas 8, 10 15 e 20 µg da proteína padrão BSA (Bovine Serum Albumin) (0,1%), para que uma curva padrão fosse estabelecida e permitisse a quantificação relativa do teor de proteína das amostras. As placas eram colocadas em banho-maria à 37°C, por 30 minutos, e após resfriar em temperatura ambiente por 20 minutos, eram obtidas as absorvâncias em 560 nm.

Determinação do teor de proteína na geração F₃ (Kjeldahl).

Para confirmar o teor de proteínas das plantas F2 selecionadas, eram analisadas amostras de sementes das plantas F3 geradas de cada semente F2 selecionada. Neste caso, a determinação de proteínas foi efetuada segundo o método Kjeldahl, para a quantificação de nitrogênio total descrito pela AOAC (1975), com modificações. Na fase de digestão, após obtenção do material digerido, adicionou-se peróxido de hidrogênio 30%, levando a mistura a aquecimento por mais 30 minutos. Na fase de destilação, recolheu-se a amônia liberada em solução de ácido bórico 4%. Obteve-se o teor de nitrogênio pela titulação da amônia com ácido clorídrico 0,05 N. A partir do teor de nitrogênio, foi calculada a porcentagem de proteína total da amostra, empregando-se o fator de 6,25. Os resultados dos teores de proteína foram expressos em porcentagem com base na matéria seca, constituindo a média de três repetições.

Extração de DNA e amplificação

O DNA das plantas foi extraído pelo método descrito por DOYLE e DOYLE (1990) com modificações. As reações de amplificação de microssatélites foram feitas em volume total de 15 μ L, contendo 12,5 mM de Tris-HCl (pH 8,3), 62,5 mM de KCl, 2,5 mM de $MgCl_2$, 125 μ M de cada um dos deoxinucleotídios (dATP, dTTP, dGTP e dCTP), 0,6 μ M de cada *primer*, uma unidade da enzima Taq DNA-polimerase e 18 ng de DNA. As amplificações foram realizadas em termociclador Perkin Elmer 9600 programado para uma etapa inicial de 7 minutos a 72°C, seguida de 30 ciclos de 1 minuto a 94°C, 1 minuto a 50°C e 2 minutos a 72°C. Por fim, uma etapa de 7 minutos a 72°C. Os fragmentos amplificados eram separados por eletroforese em gel não desnaturante de poliacrilamida 10% por 4 horas a 100 V, e corados com solução de brometo de etídio (0,2 μ g/mL). O gel possuía 16,5x12,5 cm e espessura de 0,5 mm. Terminada a eletroforese, os géis eram fotodigitalizados pelo sistema de fotodocumentação Eagle Eye II (Stratagene).

Confirmação do cruzamento das plantas F1

Todas as plantas F1 foram confirmadas quanto ao seu caráter híbrido com o auxílio de marcadores microssatélites. *Primers* de SSR confirmados como polimórficos no *fingerprinting* para cada cruzamento foram utilizados para detecção das plantas heterozigotas.

Escolha dos melhores cruzamentos com base na capacidade geral de combinação (CGC) e distâncias genéticas entre os progenitores.

O teor de proteína de sementes F3 derivadas de sementes F2 selecionadas foi determinado pelo método de Kjeldahl tomando-se 10 sementes de cada planta F2. Cada cruzamento teve diferente número de repetições e diferentes progenitores, como pode ser visto na Tabela 1, sendo considerado cada cruzamento um tratamento. A análise de variância

e a análise dos melhores cruzamentos pelo modelo hierárquico foram efetuadas pelo programa GENES (CRUZ 1997), como mostra o modelo a seguir:

$$\text{Modelo} = \mu + P_i + S/P_{ij} + E_{ij}$$

em que:

μ : média geral da população;

P_i : efeito do pai i ;

S/P_{ij} : efeito do pai ij dentro de fêmea;

E_{ij} : erro aleatório associado ao modelo.

As distâncias genéticas entre os progenitores foram determinadas pelo coeficiente de coincidência simples efetuado também pelo programa estatístico GENES (CRUZ 1997).

Ganhos de seleção

Os teores de proteína das sementes selecionadas na geração F2 foram confirmados na geração F3 pelo método de Kjeldahl. Os ganhos quanto à seleção em F2 foram calculados com base no teor de proteína do progenitor recorrente, como:

$$\%GS = GS * 100/\%PTN \text{ Recorrente}$$

em que:

$\%GS$ = Ganho percentual de proteína;

GS = Ganho real de proteína;

$\%PTN$ recorrente = Porcentagem de proteína do progenitor recorrente.

RESULTADOS E DISCUSSÃO

Em programas de melhoramento assistidos por marcadores moleculares pode-se ter diversos tipos de seleção baseado em marcadores. Pode-se, por exemplo, com o auxílio de marcadores moleculares, confirmar a natureza híbrida de sementes resultantes de cruzamentos entre dois progenitores, selecionar indivíduos mais próximos do progenitor recorrente e selecionar genótipos desejáveis com base em marcadores estreitamente ligados às características de interesse. No presente trabalho, os marcadores microssatélites foram utilizados para confirmar a natureza híbrida das plantas F1 e na seleção de indivíduos geneticamente mais próximos do progenitor recorrente. Não foi possível selecionar genótipos com base em marcadores, uma vez que o programa de melhoramento genético da qualidade da soja em andamento no BIOAGRO/UFV ainda não dispõe de marcadores moleculares estreitamente associados à característica alto teor de proteína.

Conforme mostrado na Figura 1, marcadores microssatélites foram utilizados na identificação da natureza híbrida das plantas F1 para confirmação dos cruzamentos. As plantas que apresentam dois alelos, heterozigotas, (duas bandas) indicam que foram obtidas por hibridação, enquanto que aquelas que apresentam somente uma banda são derivadas de autofecundação. Na Figura 1 pode-se observar que as plantas 12, 14, 15 e 29 são resultantes de autofecundação. Aquelas plantas confirmadas como híbridas foram autofecundadas e as populações F2 foram selecionadas com base no teor médio de proteína das sementes da planta F1 e na distância genética entre os progenitores que lhes deram origem. A Tabela 2 mostra as plantas F1 oriundas de todos os sete cruzamentos que foram selecionadas com base no alto teor de proteína e menores distâncias genéticas a fim de totalizar 400 sementes F2 para cada cruzamento. Pode-se verificar que o teor de proteína variou bastante, dentro e entre cruzamentos, indicando alta variabilidade genética para esta característica. O menor teor médio de proteína de planta F1 foi 42,19% para o cruzamento entre CD204 e OC671PT46-2 e o maior teor foi 47,73% para o cruzamento entre OC953312 e CD202PT31-3. O cruzamento oriundo de OC 953194 não foi avaliado pelo

método do BCA pois este material foi descartado do programa devido a características indesejáveis em performance de campo. O cruzamento derivado de CD985015 só será avaliado em F3 pois se encontra em estágio mais avançado em relação a estes outros cruzamentos.

A distância genética variou de 0,40 entre CD211 e OC671PT46-1 a 0,67 entre CD204 e CD202PT31-3. O principal objetivo desta seleção prévia foi obter populações de plantas com alto teor de proteína e com menor distância genética em relação ao progenitor recorrente. No entanto, em alguns casos ocorreram dificuldades para atender a estes dois pré-requisitos e, conseqüentemente, o teor de proteína foi priorizado em detrimento da distância genética. Este foi o caso do cruzamento envolvendo o progenitor recorrente CD204 com os vários progenitores doadores, onde as distâncias genéticas são relativamente altas ($> 0,50$). Além disto, é importante, sempre que possível, realizar seleções em populações F2 oriundas do cruzamento com mais de um progenitor doador de forma a garantir a variabilidade genética para os próximos ciclos de retrocruzamentos.

Em seguida todos os cruzamentos foram submetidos à análise de variância. Observa-se na Tabela 2 que todos os tratamentos (cruzamentos) foram significativos pelo teste F a 1 e 5 % de probabilidade, indicando variabilidade genética para teor de proteína nos vários cruzamentos. A média geral foi de 42,44 % e o coeficiente de variação (CV) foi baixo, indicando uma boa precisão experimental. O coeficiente de variação genético (CVg) igual a 3,85, a relação CVg/CV, superior à unidade, e o alto valor de herdabilidade, igual a 88,72, refletem uma situação bastante favorável à seleção para alto teor de proteína.

Na Tabela 4 foi avaliado o efeito de cada cruzamento utilizando-se o modelo hierárquico. Há variabilidade genética em todos os cruzamentos (significativos a 5% de probabilidade) e, ainda, os cruzamentos para CD204, CD210, OC953006, OC953194 e OC953312 apresentaram variabilidade genética significativa a 1 % de probabilidade pelo teste de F. As médias do teor de proteína variaram de 39,6 % para o cruzamento de OC953006 e de 44,12% para o cruzamento de OC953312. Novamente, foram observados altos valores de herdabilidade para todos os cruzamentos, indicando que a seleção para alto teor de proteína seria bem sucedida. Vários autores têm

relatado variações nos valores de herdabilidade para teor de proteína em soja. No trabalho de COBER e VOLDENG (2000) as estimativas de herdabilidade no sentido amplo para proteína variaram de 0,40 a 0,77, em linhagens avançadas derivadas de um único ciclo de retrocruzamento e de um único cruzamento. KWON e TORRIE (1964) avaliaram a herdabilidade de várias características agrônômicas, incluindo teor de proteína em uma população, encontrando para esta característica valor de herdabilidade de 57%. ERICKSON *et al.* (1981), obtiveram valores de herdabilidade para teor de proteína em F2 de 27%, estimada pela regressão pai-filho, e de 78% no sentido amplo, utilizando médias de famílias F3 cultivadas em dois locais, em uma única época de cultivo.

A relação entre CVg e CV (Tabela 4) para alguns cruzamentos foi abaixo de 1, indicando que o progresso genético nestes casos poderia ser reduzido devido a um maior efeito do ambiente no teor de proteína nas populações destes cruzamentos. Os melhores cruzamentos foram escolhidos com base na maior variância genética e maior média do teor de proteína, o que evidencia maior capacidade geral de combinação (CGC). Verifica-se, também, na Tabela 4, que os melhores cruzamentos foram aqueles que envolveram o progenitor recorrente OC953312, obtendo maior variabilidade, em torno de 4,62, maior valor de herdabilidade (93%) e a maior relação CVg/CV (1,57).

Na Tabela 5 é apresentado o teste de Tukey para as médias do teor de proteína dos 48 tratamentos (cruzamentos). Médias seguidas pela mesma letra não diferem entre si pelo teste de Tukey a 5% de probabilidade. Conforme salientado na Tabela 5, os progenitores doadores de alelos para alto teor de proteína foram selecionados com base nas maiores médias, sendo eles: OC128PT258, CD203PT30-3, CD203PT30-2, CD203PT31-3, CD202PT31-3, CD202PT31-2 e CD201PT82-3. A seleção desses progenitores doadores como sendo as melhores fontes de genes para alto teor de proteína para dar prosseguimento a este e outros programas de melhoramento reduzirá em muito a necessidade de utilização de um grande número de progenitores doadores. Conseqüentemente isto diminuirá o número de cruzamentos necessários para se obter as melhores populações

as quais terão altas médias para teor de proteína, reduzindo assim o número de análises a serem efetuadas e os custos do programa.

COBER e VOLDENG (2000) demonstraram que o esquema de melhoramento de um único cruzamento pode ser tão eficiente, no desenvolvimento de linhagens de soja com alto teor de proteína e alta produtividade, quanto o esquema de um único retrocruzamento com o progenitor recorrente de alta produtividade, desde que o progenitor de alto teor de proteína seja um germoplasma adaptado.

Quatrocentas sementes F2 de cada cruzamento foram analisadas pelo método do ácido bicinconínico (SMITH *et al.* 1985), e 10% das sementes com maiores teores relativos de proteína foram selecionadas e plantadas. A Figura 2 mostra um gráfico onde foram plotados teores de proteína da população F2 derivada dos cruzamentos que envolveram o progenitor recorrente OC953312 *versus* o número de sementes analisadas. Pode ser observada nesta Figura 2 uma distribuição normal para teor de proteína, o que é indicativo de uma característica quantitativa. ISHIGE (1984), demonstrou que existe uma influência de efeito materno no teor de proteína em soja, além de um significativo efeito aditivo. Como no trabalho, somente o efeito genético aditivo teve significância, este autor sugeriu que o teor de proteína teria um alto valor de herdabilidade. Ele também estimou o número de genes para teor de proteína, mostrando que esta característica estava sob o controle genético de 2 a 3 genes. No entanto, este autor sugere que para se fazer estimação mais precisa do número de genes é necessário que se use materiais (genótipos) o mais divergente possível.

Com a finalidade de testar a eficiência dos marcadores microssatélites, selecionados e otimizados para *fingerprinting* dos progenitores de soja usados neste trabalho na seleção assistida (RC-SAM), as distâncias genéticas, das plantas F2 selecionadas e confirmadas pelo teor de proteína na geração F3 do cruzamento envolvendo o progenitor recorrente CD985015, foram determinadas. A Tabela 5 apresenta os teores de proteína de sementes F3 derivadas das plantas F2 selecionadas, os ganhos e as distâncias genéticas (estimadas por marcadores microssatélites) em relação ao progenitor recorrente CD985015. Para estimação das distâncias genéticas entre o progenitor recorrente e as

plantas F2 selecionadas foram utilizados 19 dos 44 pares de *primers* previamente selecionados. Foram identificadas quatro plantas F2, que apresentaram em média 43,03% de proteína e, ainda, geneticamente mais próximas do progenitor recorrente. O ganho absoluto das plantas selecionadas foi de 5,68, o ganho percentual foi de 15,2% e a distância genética em relação ao progenitor recorrente foi em média 0,44. A Figura 3 mostra que não houve muita variação no teor de proteína das plantas confirmadas em F3 enquanto que houve variação em suas distâncias.

As Tabelas de 7 a 12 apresentam os teores de proteína dos progenitores recorrentes CD 204, CD 210, CD 211, OC 953006, OC 953312 e CD 983343 e das plantas F2 selecionadas, e os ganhos dessas em relação aos progenitores recorrentes. As médias do teor de proteína das plantas selecionadas F2 foram de 45,78% (CD 204), 46,17% (CD 210), 46,97% (CD 211), 45,74% (OC 953006), 45,78% (OC 953312) e 47,47 (CD 983343). Os ganhos médios reais variaram de 3,88 a 9,35 e percentuais de 9,17 a 24,86%, respectivamente.

Ao analisar as Tabelas de 7 a 12 observa-se que houve aumentos substanciais para teor de proteína em todos os cruzamentos. No entanto, o maior ganho (24,86%) foi em relação ao progenitor recorrente CD 211 que apresentou menor teor de proteína (37,62%) (Tabela 9), e o menor ganho (9,17%) foi em relação ao progenitor recorrente CD 210 que apresentou maior teor de proteína (42,29%) (Tabela 8).

As plantas de cada cruzamento que foram selecionadas para alto teor de proteína serão submetidas a *fingerprinting* de DNA com o auxílio de marcadores microssatélites para estimar as distâncias genéticas em relação ao progenitor recorrente, em que as mais próximas serão selecionadas para dar continuidade ao programa de retrocruzamentos.

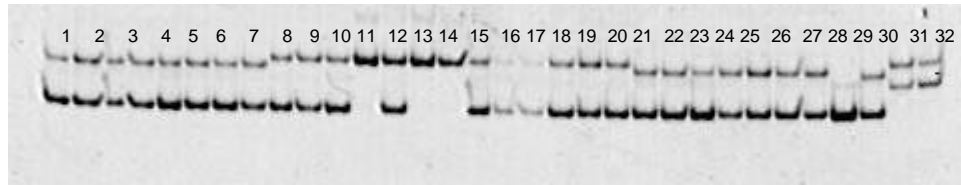


Figura 1 – Identificação da natureza híbrida das plantas F1 por marcadores microssatélites. As canaletas 12, 14, 15 e 29 mostram plantas homocigotas para um dos alelos de microssatélite e portanto são resultantes de autofecundação.

Tabela 2 – Cruzamentos escolhidos para seleção individual em F2 pelo método do ácido bicinconínico (BCA) com base no teor médio de proteína da planta F2 e na distância genética entre os progenitores.

Progenitor recorrente	Progenitor doador	Proteína F2 (%)	Distância genética
CD204	OC671PT46-2-1	42,19	0,65
	OC671PT46-2-1	42,95	0,65
	OC671PT46-2-1	45,07	0,65
	OC671PT46-2-1	44,46	0,65
	CD202PT31-3	45,27	0,67
	CD202PT31-3	42,76	0,67
	CD202PT31-3	43,29	0,67
	OC128PT258	44,79	0,64
	OC128PT258	45,61	0,64
	OC128PT258	43,7	0,64
	OC128PT258	43,23	0,64
	CD201PT82-3-15	43,78	0,58
	CD210	OC671PT46-2	43,5
OC671PT46-2		43,2	0,46
OC671PT46-2		42,4	0,46
OC671PT46-2		42,6	0,46
CD201PT82-3		44,0	0,57
CD211	CD203PT30-2	43,9	0,60
	CD203PT30-2	44,42	0,60
	OC671PT46-1	44,89	0,40
	OC671PT46-1	44,52	0,40
	OC671PT46-1	45,56	0,40
	OC671PT46-1	44,10	0,40
CD983343	OC128PT258	44,2	0,48
	OC128PT258	43,5	0,48
	OC128PT258	43,0	0,48
	OC128PT258	43,2	0,48
	OC128PT258	45,56	0,48
	OC128PT258	45,5	0,48
OC953006	CD202PT31-1	42,62	0,42
	CD202PT31-3	43,04	0,48
	OC671PT46-2	43,71	0,50
	OC671PT46-2	43,95	0,50
	CD203PT30-2	49,46	0,63
	CD202PT111-1	43,62	0,41
	CD202PT111-1	42,96	0,41
	OC14PT260	43,17	0,48
OC953312	OC14PT260	44,39	0,48
	OC671PT46-2	43,98	0,55
	CD202PT31-3	46,14	0,42
	CD202PT31-3	47,73	0,42
	CD202PT31-3	46,47	0,42
	CD202PT31-3	44,52	0,42
	CD202PT11-1	44,24	0,41
OC128PT258	44,96	0,59	

Tabela 3 - Análise de variância para os cruzamentos entre as variedades elite e as linhagens de alto teor de proteína

Tratamentos	47
QMT	16,5**
QMR	1,87
Média geral	42,44
CV (%)	3,22
CVg	3,85
σ^2_p	3,0
σ^2_g	2,65
σ^2_e	0,34
h^2	88,72

Significativo pelo teste F a 1% de probabilidade

Tabela 4 – Análise variância do modelo hierárquico avaliando o efeito dos sete cruzamentos e estimativas dos parâmetros: variância genética (σ^2_g), herbabilidade (h^2) e relação entre os coeficientes de variação genético e experimental (CV_g/CV_e).

	QM	F	P	Média	CGC	σ^2_g	h^2	CV_g/CV_e
Tratamentos	16,5	8,81**	0	42,64	0,20	2,65	0,88	1,19
CD 204	6,26	3,34**	0,00727	42,63	0,19	0,79	0,70	0,65
CD 210	6,96	3,72**	0,002993	43,10	0,66	0,92	0,73	0,70
CD 211	6,47	3,46*	0,01728	42,70	0,26	0,83	0,71	0,67
CD 983343	7,86	4,20*	0,02705	42,40	0,04	1,08	0,76	0,76
OC 953006	5,04	2,69**	0,005428	41,90	-0,54	0,57	0,63	0,55
OC 953194	15,70	8,40**	0,000302	39,60	-0,54	2,51	0,88	1,16
OC 953312	27,36	14,6**	0	44,12	1,68	4,62	0,93	1,57

**Significativo a 1% de probabilidade, *Significativo a 5% de probabilidade pelo teste F.

Tabela 5 – Teste de comparação entre as médias pelo teste de Tukey a 5 % de probabilidade dos 48 tratamentos para teor de proteína.

Tratamento	Número de repetições	Teor de proteína	
OC953312 X CD201PT82-3	4	47,10	a
OC953312 X CD203PT30-3	4	46,65	ab
OC953312 X CD203PT31-3	4	46,22	ab
OC953312 X CD203PT30-2	10	45,61	ab
OC953312 X OC128PT258-2	2	45,11	abc
CD210 X CD202PT31-3	2	44,33	abcd
CD204 X OC128PT258	4	44,33	abcd
CD210 X CD202PT31-2	8	44,22	abcd
CD211 X CD203PT30-2	2	44,16	abcd
CD983343 X OC128PT258	6	43,65	abcd
CD204 x CD203PT30-2	9	43,64	bcd
CD204 X OC13PT158-3	2	43,63	bcde
CD983343 X CD203PT30-2	4	43,61	bcde
OC953006 X CD203PT30-2	10	43,45	bcde
CD210 X CD203PT30-1	2	43,35	bcde
CD210 X OC671PT46-2	4	42,94	bcde
OC953312 X CD203PT30-1	4	42,86	bcde
OC953006 X OC671PT46-2	3	42,82	bcde
CD204 X CD203PT30-1	7	42,75	cde
CD204 X CD202PT31-3	5	42,73	cde
OC953312 X CD202PT111-1	7	42,69	cde
OC953006 X CD203PT30-3	6	42,57	cde
CD204 X CD202PT31-1	3	42,55	cde
CD211 X OC671PT46-1	15	42,54	cde
CD211 X CD203PT30-1	3	42,50	cde
CD204 X CD202PT111-1	3	42,48	cde
OC953006 X CD201PT82-3	7	42,37	cde
CD983343 X CD203PT30-1	2	42,17	cdef
OC953312 X OC671PT46-2	11	42,08	cdef
CD210 X CD201PT82-3	2	42,05	cdef
OC953006 X OC14PT260	7	41,95	cdef
CD983343 X OC671PT46-1	6	41,92	cdef
CD204 X OC671PT46-2	6	41,83	cdef
OC953312 X OC671PT46-1	6	41,58	cdef
CD210 X OC671PT100-1	2	41,58	cdef
OC953006 X CD202PT31-3	4	41,56	cdef
CD211 X CD202PT111-1	3	41,55	cdef
CD204 X OC671PT46-1	5	41,54	cde
OC953006 X CD202PT31-1	4	41,48	cdef
OC953006 X CD202PT111-1	6	41,47	cdef
OC953312 X OC13PT158-3	14	41,36	cdef
OC953194 X OC671PT46-2	14	41,36	cdef
CD983343 X CD202PT111-1	2	40,86	cdef
CD204 X OC671PT100-1	6	40,79	cdef
OC953006 X OC671PT46-1	10	40,72	def
OC953006 X CD202PT31-2	2	40,34	def
OC953194 X CD203PT30-2	6	39,57	ef
OC953194 X OC671PT46-1	9	37,98	f

BCA F2 OC953312

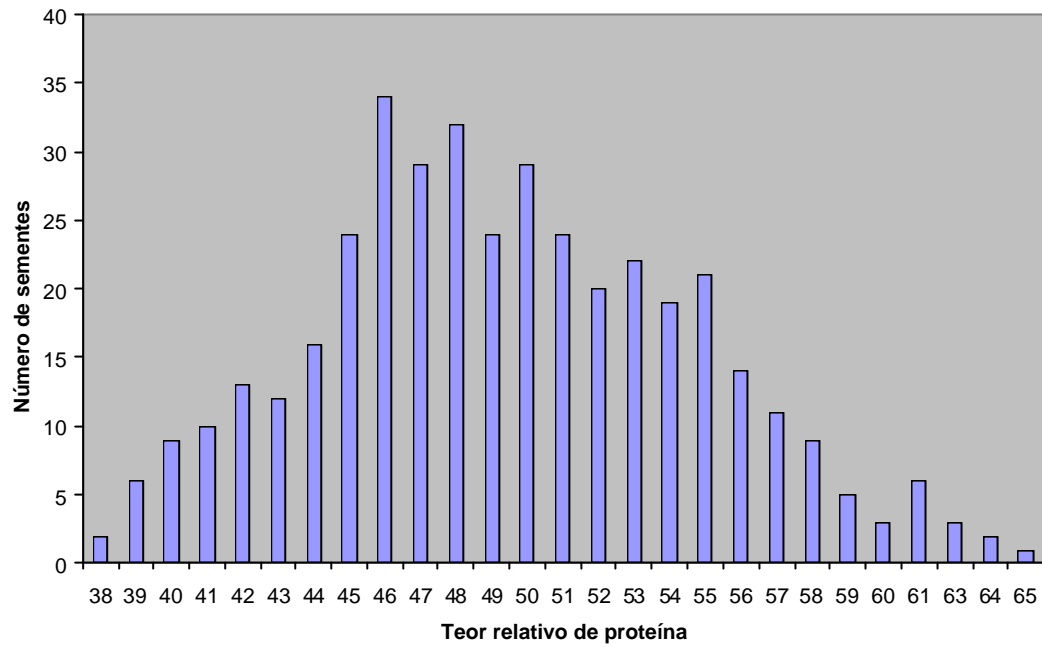


Figura 2 – Distribuição do conteúdo de proteína nas populações F2 dos cruzamentos entre o recorrente OC953312 e os doadores OC671PT46-2, CD202PT31-1, CD202PT111-1 e OC128PT258-2.

Tabela 6 - Teor de proteína do progenitor recorrente CD985015, da geração F3 derivada de sementes F2 selecionadas, ganhos em relação ao progenitor recorrente e distância genética das plantas F2 selecionadas em relação ao progenitor recorrente.

Genótipos	% Proteína			
CD 985015	37,38	Ganho ²	Ganho ³ (%)	Distância genética
CD 9850151PTA1	43,21	5,83	15,60	0,44
CD 9850151PTA8	42,28	4,90	13,10	0,50
CD 9850151PTA41	42,54	5,16	13,80	0,47
CD 9850151PTA53	42,97	5,59	14,95	0,67
CD 9850151PTA130	42,52	5,14	13,75	0,67
CD 9850151PTA252	43,71	6,33	16,93	0,39
CD 9850151PTA300	42,78	5,40	14,45	0,47
Média ¹	43,03¹	5,68²	15,20³	0,44

¹Média dos indivíduos selecionados

²Ganho absoluto em relação ao progenitor recorrente.

³Ganho percentual em relação ao progenitor recorrente.

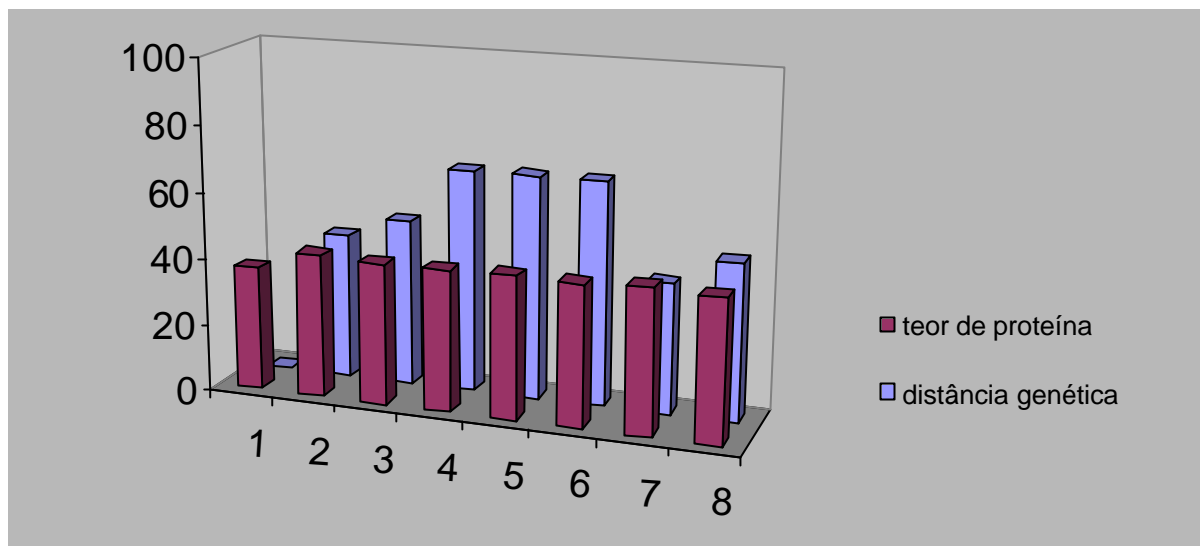


Figura 3 – Variação das distâncias genéticas dos indivíduos selecionados para alto teor de proteína em relação ao progenitor recorrente CD 985015.

Tabela 7 – Teor de proteína do progenitor recorrente CD204, da geração F3 derivada de sementes F2 selecionadas, e ganhos reais e percentuais em relação ao progenitor recorrente.

Genótipos	% Proteína	Ganho ²	Ganho ³ (%)
CD204	37,37		
CD2041PTA111	44,49	7,12	19,05
CD2041PTA360	47,03	9,66	25,85
CD2041PTA363	45,74	8,37	22,40
CD2041PTA392	45,22	7,85	21,01
CD2041PTA10	45,23	7,86	21,03
CD2041PTA170	48,96	11,59	31,01
CD2041PTA203	44,37	7,00	18,73
CD2041PTA280	44,95	7,58	20,28
CD2041PTA288	47,96	10,59	28,34
CD2041PTA361	44,61	7,24	19,37
CD2041PTA385	45,29	7,92	21,19
CD2041PTA391	48,26	10,89	29,14
CD2041PTA348	45,27	7,90	21,14
CD2041PTA330	44,88	7,51	20,10
CD2041PTA46	44,94	7,57	20,26
CD2041PTA16	45,90	8,53	22,83
CD2041PTA155	48,08	10,71	28,66
CD2041PTA184	44,97	7,60	20,34
CD2041PTA193	46,62	9,25	24,75
CD2041PTA245	44,16	6,79	18,17
CD2041PTA308	44,40	7,03	18,81
Média ¹	45,78	8,41	22,50

¹Média dos indivíduos selecionados

²Ganho absoluto em relação ao progenitor recorrente

³Ganho percentual em relação ao progenitor recorrente

Tabela 8 – Teor de proteína do progenitor recorrente CD210, da geração F3 derivada de sementes F2 selecionadas, e ganhos reais e percentuais em relação ao progenitor recorrente

Genótipos	% Proteína	Ganho ²	Ganho ³ (%)
CD210	42,29		
CD2101PTA273	44,06	1,77	4,19
CD2101PTA367	45,63	3,34	7,90
CD2101PTA432	45,43	3,14	7,42
CD2101PTA490	44,32	2,03	4,80
CD2101PTA39	47,11	4,82	11,40
CD2101PTA153	44,4	2,11	4,99
CD2101PTA188	47,64	5,35	12,65
CD2101PTA292	45,19	2,90	6,86
CD2101PTA286	48,63	6,34	14,99
CD2101PTA113	45,64	3,35	7,92
CD2101PTA131	49,29	7,00	16,55
CD2101PTA192	44,82	2,53	5,98
CD2101PTA246	47,59	5,30	12,53
CD2101PTA216	44,25	1,96	4,63
CD2101PTA365	47,42	5,13	12,13
CD2101PTA412	46,62	4,33	10,24
CD2101PTA433	48,65	6,36	15,04
CD2101PTA453	46,24	3,95	9,34
CD2101PTA486	45,91	3,62	8,56
CD2101PTA339	44,54	2,25	5,32
Média ¹	46,17	3,88	9,17

¹Média dos indivíduos selecionados

²Ganho absoluto em relação ao progenitor recorrente

³Ganho percentual em relação ao progenitor recorrente

Tabela 9 – Teor de proteína do progenitor recorrente CD211, da geração F3 derivada de sementes F2 selecionadas, e ganhos reais e percentuais em relação ao progenitor recorrente

Genótipos	% Proteína	Ganho ²	Ganho ³ (%)
CD211	37,62		
CD2111PTA306	45,61	7,99	21,24
CD2111PTA9	47,82	10,2	27,11
CD2111PTA16	46,23	8,61	22,89
CD2111PTA24	45,07	7,45	19,80
CD2111PTA36	47,99	10,37	27,57
CD2111PTA47	45,97	8,35	22,20
CD2111PTA43	46,55	8,93	23,74
CD2111PTA84	44,03	6,41	17,04
CD2111PTA91	49,58	11,96	31,79
CD2111PTA98	44,80	7,18	19,09
CD2111PTA118	47,35	9,73	25,86
CD2111PTA148	51,19	13,57	36,07
CD2111PTA140	44,24	6,62	17,60
CD2111PTA153	45,10	7,48	19,88
CD2111PTA205	44,83	7,21	19,17
CD2111PTA232	48,21	10,59	28,15
CD2111PTA245	47,87	10,25	27,25
CD2111PTA250	44,84	7,22	19,19
CD2111PTA256	45,53	7,91	21,03
CD2111PTA288	45,16	7,54	20,04
CD2111PTA307	46,59	8,97	23,84
CD2111PTA316	47,83	10,21	27,14
CD2111PTA332	51,02	13,40	35,62
CD2111PTA330	44,86	7,24	19,25
CD2111PTA340	49,95	12,33	32,78
CD2111PTA396	54,20	16,58	44,07
CD2111PTA382	45,51	7,89	20,97
CD2111PTA102	47,27	9,65	25,65
Média ¹	46,97	9,35	24,86

¹Média dos indivíduos selecionados

²Ganho absoluto em relação ao progenitor recorrente

³Ganho percentual em relação ao progenitor recorrente

Tabela 10 – Teor de proteína do progenitor recorrente OC 953006, da geração F3 derivada de sementes F2 selecionadas, e ganhos reais e percentuais em relação ao progenitor recorrente

Genótipos	% Proteína	Ganho ²	Ganho ³ (%)
OC 953006	39,16		
OC 9530061PTA44	46,38	7,22	18,44
OC 9530061PTA103	44,28	5,12	13,07
OC 9530061PTA108	46,08	6,92	17,67
OC 9530061PTA137	44,24	5,08	12,97
OC 9530061PTA157	45,61	6,45	16,47
OC 9530061PTA217	44,10	4,94	12,61
OC 9530061PTA240	45,71	6,55	16,73
OC 9530061PTA264	45,25	6,09	15,55
OC 9530061PTA296	45,11	5,95	15,19
OC 9530061PTA355	44,98	5,82	14,86
OC 9530061PTA5	44,16	5,00	12,77
OC 9530061PTA10	46,14	6,98	17,82
OC 9530061PTA39	46,57	7,41	18,92
OC 9530061PTA51	45,86	6,70	17,11
OC 9530061PTA71	47,85	8,69	22,19
OC 9530061PTA86	47,41	8,25	21,07
OC 9530061PTA262	47,38	8,22	20,99
OC 9530061PTA366	46,21	7,05	18,00
Média ¹	45,74	6,58	16,80

¹Média dos indivíduos selecionados

²Ganho absoluto em relação ao progenitor recorrente

³Ganho percentual em relação ao progenitor recorrente

Tabela 11 – Teor de proteína do progenitor recorrente OC 953312, da geração F3 derivada de sementes F2 selecionadas, e ganhos reais e percentuais em relação ao progenitor recorrente

Genótipos	% Proteína	Ganho ²	Ganho ³ (%)
OC 953312	39,03		
OC 9533121PTA11	44,57	5,54	14,19
OC 9533121PTA82	48,44	9,41	24,11
OC 9533121PTA141	45,45	6,42	16,45
OC 9533121PTA155	44,40	5,37	13,76
OC 9533121PTA158	47,15	8,12	20,80
OC 9533121PTA164	44,46	5,43	13,91
OC 9533121PTA193	46,34	7,31	18,73
OC 9533121PTA242	44,41	5,38	13,78
OC 9533121PTA308	45,17	6,14	15,73
OC 9533121PTA332	45,67	6,64	17,01
OC 9533121PTA363	44,52	5,49	14,07
OC 9533121PTA371	44,77	5,74	14,71
OC 9533121PTA389	45,22	6,19	15,86
OC 9533121PTA392	45,28	6,25	16,01
OC 9533121PTA356	47,83	8,80	22,55
OC 9533121PTA62	46,72	7,69	19,70
OC 9533121PTA67	47,92	8,89	22,78
Média ¹	45,78	6,75	17,30

¹Média dos indivíduos selecionados

²Ganho absoluto em relação ao progenitor recorrente

³Ganho percentual em relação ao progenitor recorrente

Tabela 12 – Teor de proteína do progenitor recorrente CD 983343, da geração F3 derivada de sementes F2 selecionadas, e ganhos reais e percentuais em relação ao progenitor recorrente.

Genótipos	% Proteína	Ganho ²	Ganho ³ (%)
CD983343	38,6		
CD9833431PTA267	47,04	8,44	21,87
CD9833431PTA192	44,61	6,01	15,57
CD9833431PTA332	45,41	6,81	17,64
CD9833431PTA2	50,83	12,23	31,68
CD9833431PTA17	47,85	9,25	23,96
CD9833431PTA49	44,17	5,57	14,43
CD9833431PTA42	46,43	7,83	20,28
CD9833431PTA72	45,63	7,03	18,21
CD9833431PTA59	44,21	5,61	14,53
CD9833431PTA104	46,56	7,96	20,62
CD9833431PTA193	47,65	9,05	23,45
CD9833431PTA116	51,27	12,67	32,82
CD9833431PTA143	49,36	10,76	27,88
CD9833431PTA266	46,08	7,48	19,38
CD9833431PTA287	47,58	8,98	23,26
CD9833431PTA279	44,95	6,35	16,45
CD9833431PTA307	46,51	7,91	20,49
CD9833431PTA315	47,76	9,16	23,73
CD9833431PTA348	47,70	9,10	23,58
CD9833431PTA333	44,56	5,96	15,44
CD9833431PTA380	46,14	7,54	19,53
CD9833431PTA387	48,05	9,45	24,48
CD9833431PTA168	52,21	13,61	35,26
CD9833431PTA182	51,26	12,66	32,80
CD9833431PTA208	49,45	10,85	28,11
CD9833431PTA249	48,09	9,49	24,59
CD9833431PTA366	50,27	11,67	30,23
Média ¹	47,47	8,87	22,97

¹Média dos indivíduos selecionados

²Ganho absoluto em relação ao progenitor recorrente

³Ganho percentual em relação ao progenitor recorrente

REFERÊNCIAS BIBLIOGRÁFICAS

- ASSOCIATION OF OFFICIAL ANALYTICAL CHEMISTS – AOAC. **Official methods of analysis**. Washington, D.C.: 1975. 1094p.
- CRUZ, C.D. Programa genes aplicativo computacional em genética e estatística. **Viçosa, MG: UFV**, 1997. 63p.
- COBER, E.R., VOLDENG, H.D., Developing high protein, high-yield soybean populations and lines. **Crop Sci.** v.40, n.1, p.39-42, 2000.
- DOYLE, J.J, DOYLE, J.L. Isolation of plant DNA from fresh tissue. **Focus**, v.12, p.13-15, 1990.
- ERICKSON, L.R., VOLDENG, H.D., BEVERSDORF, W.D. Early generation selection for protein in *Glycine max* X *Glycine soja* crosses. **Can. J. Plant Sci.** vol.61, p.901-908, 1981.
- FEHR, W.R., WEBER, C.R. Mass selection by seed size and specific gravity in soybean populations. **Crop Sci.**, v.8, p.551-554, 1968.
- MILLER, J.E., FEHR, W.R. Direct and indirect recurrent selection for protein in soybeans. **Crop Sci.**, v.19, p.101-106, 1979.
- HARTWIG, E.E. Breeding soybean for high seed yield and seed protein. **In: Genetic Improvement. Proceedings: World Soybean Research Conference V**. Kasetsart University Press and WSRCV, p74-83, 1994.
- KWON, S.H., TORRIE, J.H. Heritability of and interrelationships among traits of two soybean populations. **Crop Sci.**, v. 12, n.2, p.196-198, 1964.
- ISHIGE, T. Biometrical analysis and estimation of the number of genes for seed protein content of soybean *Glycine max* (L.) Merril. **JARQ**, v.17, n. 4, p.230-235, 1984.
- SMITH, P.K.; KROHN, R.I.; HERMANSON, G.T.; MALLIA, A. K.; GARTNER, F. H.; PROVENZANO, M.D.; FUJIMOTO, E.K.; GOEKE, M.M.; OLSON, B.J. KLENK, D. C. Measurement of protein using bicinconinic acid. **Anal. Biochem.**, 15: 76-85. 1985.
- WILCOX, J.R., CAVINS, J.F., Backcrossing high protein to a soybean cultivar. **Crop Sci.** vol. 35, p.1036-1041, 1995.

CAPÍTULO 4

CARACTERIZAÇÃO BIOQUÍMICA DE LINHAGENS DE SOJA COM ALTO TEOR DE PROTEÍNA

RESUMO

Os teores de proteína e óleo em sementes de soja determinam o seu valor comercial. O farelo de soja destinado à exportação é classificado em três categorias, de acordo com o seu conteúdo de proteína: HyPro (> 48%), “Normal” (46%) e LowPro (< 43,5%). Para atingir o nível classificado como “Normal” e HyPro a soja deve conter acima de 41,5% e 43% de proteína nas sementes, respectivamente, com base na matéria seca. Um dos objetivos do melhoramento da soja é obter sementes com elevado teor de proteína. Porém, faz-se necessário o conhecimento de como a composição bioquímica da soja é afetada devido ao melhoramento genético, sendo principalmente importante que os teores de aminoácidos limitantes não sejam ainda mais reduzidos. A proteína da soja é deficiente nos aminoácidos metionina e cisteína e deve ser suplementada com outras fontes de proteína, ou com metionina sintética, quando a soja é utilizada como fonte de proteína para humanos e animais monogástricos. O objetivo desse trabalho foi caracterizar isolinhas de soja com alto teor de proteína com relação aos teores de proteína, óleo, carboidratos totais, composição aminoacídica e em relação à proporção das proteínas 7S e 11S. Os materiais analisados foram duas isolinhas com alto teor de proteína e uma isolinha com ausência de lipoxigenases (UFV 105) e teor normal de proteína, todas derivadas da variedade comercial CAC –1. Os teores de proteína foram 40,68% para UFV 105, 47,78% (aumento de 17,5%) para a isolinha 1 e 46,56% (aumento de 14,5%) para a isolinha 2. Os teores de óleo foram 24,0%, 19,22% e 18,56% para a linhagem UFV-105, isolinha 1 e 2, respectivamente. Os teores de carboidratos totais foram de 30,37% para UFV 105, 27,61% e 29,22% para isolinhas 1 e 2, respectivamente. Estes resultados indicam que o aumento do teor de proteína nas isolinhas foi acompanhado por uma redução no teor de óleo e no teor de carboidratos totais. Com relação à composição aminoacídica, observou-se que o melhoramento para teor de proteína promoveu um aumento geral na quantidade de todos aminoácidos nas isolinhas 1 e 2, quando comparadas com a linhagem UFV 105, no entanto a relação S/N não foi alterada. Isso indica que o melhoramento para teor de

proteína destas linhagens derivadas de CAC-1 não afetou o valor nutricional da proteína da soja. A quantificação das proteínas de reserva 7S e 11S em gel SDS-PAGE por densitometria, mostrou que o aumento do teor de proteína observado nas isolinhas foi proporcionado, principalmente, pelo aumento nos teores da proteína 11S. Além disto, uma pequena redução foi observada nos teores da proteína 7S nas isolinhas de alto teor de proteína, indicando que poderia haver uma melhoria na qualidade nutricional do farelo de soja derivada destas isolinhas, uma vez que a proteína 11S tem melhor qualidade nutricional e melhor digestibilidade em relação à proteína 7S.

Palavras chave: composição aminoacídica, metionina, cisteína, proteínas de reserva, 7S e 11S.

INTRODUÇÃO

Sementes de variedades cultivadas de soja contém cerca de 40% de proteína e 20% de óleo com base na matéria seca. Várias proteínas diferentes são encontradas na semente de soja, mas somente poucas estão presentes em grandes quantidades. As proteínas presentes em maior quantidade são as proteínas de reserva glicinina e β -conglucina, lipoxigenases, inibidor de tripsina Kunitz (KTI), inibidores de protease de baixo peso molecular (dos quais o melhor estudado é o inibidor Bowman-Birk), lectina e urease. Destas, glicinina e β -conglucina são as proteínas predominantes, perfazendo um total de cerca de 70% das proteínas da semente (HILL & BREIDENBACH, 1974). As outras, as quais estão presentes em altos níveis comparadas com outras proteínas metabólicas perfazem, geralmente, cerca de 2 a 5% do total de proteína. As proteínas de reserva da soja têm coeficientes de sedimentação de, aproximadamente, 7S (composta pelas subunidades α' , α e β) e 11S (composta pelas subunidades A1aB2, A1bB1b, A2B1a, A3B4 e A5A4B3) (NIELSEN, 1985).

OGAWA *et al.* (1989) e KITAMURA *et al.* (1981) analisaram linhagens com variações nos teores da proteína 7S e concluíram que a diminuição na síntese da proteína 7S pode ser compensada por um aumento na síntese e deposição da proteína 11S (que tem maior conteúdo de metionina), mantendo assim o teor de proteína total não alterado. Conseqüentemente, isto aumentaria o conteúdo de aminoácidos sulfurados da proteína de soja. De fato, KITAMURA (1995) demonstrou que uma linhagem mutante com redução no conteúdo de 7S, apresentou aumento de cerca de 15% na proteína 11S em relação às variedades normais, sem contudo alterar o teor de proteína total. Neste mesmo trabalho também é mostrado que linhagens mutantes para baixo teor da proteína 11S não sofreram redução no teor da proteína 7S.

YAKLICH (2001) investigou a variação dos teores das proteínas de reserva β -conglucina e glicinina em linhagens com alto teor de proteína em comparação com linhagens com teor normal de proteína. As subunidades componentes destas proteínas foram quantificadas por densitometria. Como

conclusão, este autor relata que as linhagens de alto teor de proteína parecem conter maiores teores de β -conglucina e glicina do que as linhagens normais e, que as quantidades relativas destas duas proteínas e de seus polipeptídeos constituintes variaram entre as linhagens.

WILCOX e SHIBLES (2001) avaliaram linhagens com diferentes teores de proteína. O aumento no teor de proteína ocorreu às custas do teor de óleo, carboidratos totais e sacarose com coeficientes de correlação de -0,88, -0,71 e -0,66, respectivamente, além de aumento no teor de aminoácidos sulfurados. No entanto, a proporção S/N (enxofre/nitrogênio) se manteve constante entre os vários teores de proteína na população. Estes autores sugerem que a diminuição no teor de carboidratos totais e, principalmente, dos oligossacarídeos rafinose e estaquiose (que são fatores antinutricionais) com o aumento no teor de proteína, poderiam melhorar o valor nutricional do farelo de soja. IMSANDE (2001) identificou linhagens mutantes que produzem alto teor de metionina (18% a mais que o normal) e um teor normal de nitrogênio. Estas linhagens foram cruzadas com genótipos normais e a progênie mostrou proporção de S/N 16,2% maior que o normal. Análises de aminoácidos das sementes do progenitor recorrente e das linhagens derivadas revelou 1,51 e 1,84 mol%, respectivamente, para metionina e 1,32 e 1,685 mol%, respectivamente, para cisteína. Estes autores comentam que estas linhagens poderiam ser cruzadas com genótipos normais visando complementar os requerimentos nutricionais da soja em relação aos teores de aminoácidos sulfurados sem alterar o teor de proteína.

Isto sugere que uma avaliação geral de um genótipo é extremamente importante no final de um programa de melhoramento de forma que a modificação genética sofrida pela linhagem não altere seu valor nutricional com o melhoramento.

O objetivo desse trabalho foi o de caracterizar linhagens de soja com alto teor de proteína, obtidas no programa de melhoramento da qualidade da soja para a agroindústria do BIOAGRO/UFV, com relação ao teor de proteína, óleo, carboidratos totais, composição aminoacídica e proporção entre as proteínas de reserva 7S e 11S.

MATERIAL E MÉTODOS

Materiais Genéticos

Duas isolinhas de alto teor de proteína obtidas no programa de melhoramento genético da soja para a agroindústria do BIOAGRO/UFV e a linhagem UFV 105 (com ausência de lipoxigenases e teor normal de proteína), todas derivadas da variedade comercial CAC-1 foram avaliadas quanto à composição bioquímica de seus grãos.

Determinação de Proteínas

A determinação de proteínas foi efetuada pelo método de KJELDAHL para a quantificação de nitrogênio total descrito pela AOAC (1975), com modificações. Na fase de digestão, após obtenção do material aparentemente digerido, adicionou-se peróxido de hidrogênio 30%, levando a mistura a aquecimento por mais 30 minutos. Na fase de destilação, recolheu-se a amônia liberada em solução de ácido bórico 4%. Obteve-se o teor de nitrogênio pela titulação da amônia com ácido clorídrico 0,05 N. A partir do teor de nitrogênio, foi calculada a porcentagem de proteína total da amostra, empregando-se o fator de 6,25. Os resultados dos teores de proteína foram expressos em porcentagem com base na matéria seca, constituindo a média de três repetições.

Determinação de Lipídios

Determinaram-se lipídios em aparelho extrator de Soxhlet, utilizando éter de petróleo como solvente, segundo o procedimento descrito em Normas Analíticas do Instituto Adolfo Lutz (INSTITUTO ADOLFO LUTZ 1985), com refluxo por 10 horas.

Eletroforese em gel de poliacrilamida

As proteínas de reserva de sementes de cada genótipo foram extraídas de amostras de fragmentos dos cotilédones (10 mg) com 900 µL de tampão de extração Tris-HCl 50 mM, pH 8,0 contendo 0,2% SDS, 10 mM de β-mercaptoetanol e 6 M de uréia. As amostras foram extraídas em gral e pistilo em temperatura ambiente sendo centrifugadas por 15 min a 14.000 RPM em microcentrífuga Eppendorf modelo 5415C. Em seguida, as amostras foram desnaturadas por 3 min a 100 °C. O gel empilhador continha poliacrilamida na concentração de 6% em Tris-HCl 0,45 M, uréia 1,78 M e SDS 0,09%, pH 8,8. As proteínas nos extratos foram separadas em géis gradientes de poliacrilamida de 10 a 18 % (FONTES *et al.* 1994) contendo Tris-HCl 0,45 M, Uréia 1,78 M e SDS 0,09%, pH 8,8. Amostras de 20 µL foram aplicadas no gel e a corrida foi realizada por 80 V por uma hora e a 100 V por 5 horas. As bandas protéicas foram reveladas em solução de Coomassie Blue R-250.

Determinação do Conteúdo de Cinzas

A determinação do conteúdo de cinzas foi conduzida por meio da calcinação das amostras a 550°C, conforme o método descrito em Normas Analíticas do Instituto Adolfo Lutz (INSTITUTO ADOLFO LUTZ, 1985).

Determinação do Conteúdo de Carboidratos Totais

O teor de carboidratos totais foi obtido por diferença percentual como mostrado a seguir:

$$\%CHO = 100 - (\%PTN + \% \text{ cinzas} + \% \text{ óleo})$$

onde:

%CHO = Teor de carboidratos totais em porcentagem;

%PTN = Teor de proteína em porcentagem;

% cinzas = Teor de cinzas;

% óleo = Teor de óleo.

Determinação da composição aminoacídica

De cada genótipo foram analisadas duas repetições de amostras de sementes de soja moídas em moinho do tipo “Wiley”. A farinha obtida foi desengordurada em extrator Soxhlet. Os aminoácidos foram analisados no Centro Interdepartamental de Química de Proteínas da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo. A análise de aminoácidos foi realizada em duplicata em amostras previamente desengorduradas que sofreram hidrólise ácida para se quantificar os aminoácidos: lisina, histidina, arginina, ácido aspártico, treonina, serina, ácido glutâmico, prolina, glicina, alanina, cisteína, valina, metionina, isoleucina, leucina, tirosina e fenilalanina. Alíquotas de amostras contendo cerca de 5-10 mg de sólidos, previamente desengordurados, foram transferidas para ampolas de borossilicato de 10x150 mm (Pyrex[®]) que foram previamente pirolisadas a 400°C por 8 horas. Foram adicionados 0,5 mL de solução de ácido clorídrico 6N, bidestilado a 104°C, contendo 0,1% de fenol (m/v). Cada ampola contendo amostra e HCl foi selada, sendo anteriormente feito vácuo na mesma. Estas ampolas foram colocadas em uma mufla a 110°C, por 22 horas (MOORE *et al.* 1958). Após a hidrólise ácida em solução, a amostra foi seca em concentrador rotatório e ressuspendida em solução tampão de citrato de sódio 0,17 M, pH 2,2, contendo polietilenoglicol 400 a 15% (v/v) e tioglicol 0,4% (v/v).

A composição qualitativa e quantitativa dos aminoácidos foi determinada por análise automática, realizada por meio de cromatografia de troca iônica com derivação pós-cromatográfica com ninidrina, pelo método de SPACKMAN *et al.* (1963), utilizando-se um analisador automático (ALONSO & HIRS 1968) – construído no Centro de Química de Proteínas da FMRP/USP, Ribeirão Preto, SP. Alíquotas entre 10 e 900 µL eram aplicadas nas colunas de troca catiônica (Resina: PC - 6A Amino acid Analysis Resin - Pierce) do analisador (coluna curta para aminoácidos básicos e triptofano, e longa para aminoácidos ácidos e neutros) e eluídas por modificações de pH e força iônica (coluna curta pH 5,28; coluna longa pH 3,25 e posteriormente 4,25) (MOORE *et al.* 1958). Após a separação cromatográfica, os

aminoácidos eluídos da coluna reagiram com ninidrina a uma temperatura de aproximadamente 100°C (banho de água em ebulição) por 15 minutos e os produtos desta reação foram detectados colorimetricamente em dois comprimentos de onda: 440 nm para a Prolina e 570 nm para os demais aminoácidos. A identificação dos picos era realizada com base nos tempos de retenção de cada resíduo. O aparelho foi padronizado com uma solução padrão de aminoácidos que contém 40 nmoles de cada resíduo injetado. A altura dos picos foi utilizada para determinar o fator de cálculo, assumindo-se que cada pico apresenta uma forma gaussiana. Estas medidas são realizadas manualmente (ALONSO & HIRS 1968) e os dados de linha de base e altura máxima foram processados por uma planilha desenvolvida no próprio CQP-FMRP/USP. Este aparelho foi otimizado para operar em uma faixa linear de 0,25 a 1,20 vezes a quantidade do padrão. Diferenças entre duplicatas devem ser menores que 8%.

Composição da proteína de soja nas frações 7S e 11S

As subunidades protéicas reveladas em gel SDS – PAGE em gradiente de poliacrilamida de 10 a 18% foram analisadas por densitometria em escaner “Personal Densitometer SI” da Molecular Dynamics. As subunidades das proteínas de reserva 7S e 11S de sementes de soja foram identificadas de acordo FONTES *et al* 1984). Foram feitas duas repetições de cada amostra, na qual a quantidade relativa de proteína correspondente a cada banda foi calculada em relação a uma curva padrão estabelecida com quantidades crescentes da proteína padrão anidrase carbônica, 2, 4, 8 e 10 µg.

RESULTADOS E DISCUSSÃO

Os resultados obtidos quanto ao teor de proteína foram de 40,68% para a linhagem UFV 105, 47,8% (aumento de 17,4%) para a isolinha 1 e 46,56% (aumento de 14,5%) para a isolinha 2 (Tabela 1). Os teores de óleos variaram de 24,03%, 19,22% e 18,56%, para a linhagem UFV 105, isolinha 1 e isolinha 2, respectivamente. Como pode ser observado à medida que se aumenta o teor de proteína o teor de óleo é reduzido, comprovando a correlação negativa existente entre estas duas características. Vários autores reportam em seus trabalhos a correlação negativa existente entre proteína e óleo. Desta forma, HELMS e ORF (1998) mostraram que a seleção direta para teor de proteína produziu uma redução no teor de óleo. WILCOX (1998), realizando seleção recorrente para teor de proteína, mostrou que a seleção recorrente aumentou a média do teor de proteína nas sementes de 5,8 g kg⁻¹ e diminuiu a média do teor de óleo de 2,3 g kg⁻¹ por ciclo.

O teor de carboidratos foi determinado por diferença entre o total e o teor de proteína, óleo e cinzas. Desta forma, o teor de carboidratos foi de 30,37%, 27,61% e 29,22% para UFV-105, isolinha 1 e isolinha 2, respectivamente. Como pode ser observado, quando se aumentou o teor de proteína o teor de carboidratos também foi reduzido, o que evidencia uma associação negativa entre estas duas características. HARTWIG *et al.* (1997) sugerem que o aumento no teor de proteína poderia reduzir o teor de açúcares solúveis presentes na fração carboidratos totais, melhorando desta forma a qualidade nutricional do farelo de soja. WILCOX e SHIBLES (2001) mostraram que o aumento no teor de proteína em linhagens de soja ocorreu às custas do teor de óleo, carboidratos totais e sacarose com coeficientes de correlação de -0,88, -0,71 e -0,66, respectivamente. Estes autores concluíram que a diminuição no teor de carboidratos com o aumento do teor proteína deve contribuir para um aumento do valor nutricional do farelo obtido das linhagens melhoradas.

A Tabela 2 mostra a composição aminoacídica da linhagem UFV-105 e das isolinhas 1 e 2. Observa-se, de maneira geral, quando se considera a

porcentagem de cada aminoácido na proteína da soja, que o aumento do teor de proteína detectado nas linhagens melhoradas (isolinhas 1 e 2) não promoveu alteração perceptível e significativa na composição aminoacídica (% de cada aminoácido na proteína). Neste particular, os valores determinados para metionina foram 1,31, 1,31 e 1,30% para a linhagem UFV 105, isolinhas 1 e 2, respectivamente. Para cisteína, estes valores foram de 1,03, 0,97 e 1,06%, para a linhagem UFV 105, isolinhas 1 e 2, respectivamente. Portanto, os teores dos aminoácidos sulfurados metionina e cisteína, que são considerados os aminoácidos mais limitantes da proteína da soja, não são diferentes entre a linhagem UFV 105 e as isolinhas com alto teor de proteína. Isto indica que o aumento observado no teor de proteína nas isolinhas melhoradas não promoveu alteração perceptível no valor nutricional da proteína de soja.

A análise dos polipetídeos constituintes das proteínas de reserva da soja no gel SDS-PAGE em gradiente de poliacrilamida de 10 a 18% contendo uréia 6 M (Figura 1), evidencia um aumento nas quantidades das subunidades componentes da proteína glicínica nas isolinhas de alto teor de proteína. As canaletas 1, 2, 3 e 4 indicam quantidades crescentes do padrão anidrase carbônica, ou seja, 2, 4, 8 e 10 micrograma de proteína, respectivamente. As canaletas 5 a 13 mostram a linhagem UFV 105 e as isolinhas 1 e 2 em três triplicatas.

Os resultados da quantificação do gel por densitometria (Tabelas 2 e 3) mostram que a quantidade da subunidade 7S foi maior na linhagem UFV 105 que possui teor normal de proteína e menor nas isolinhas com alto teor de proteína. Por outro lado, a quantidade da proteína 11S foi maior nas isolinhas de alto teor de proteína. A Tabela 2 apresenta os valores de cada subunidade, mostrando que a redução da proteína 7S foi proporcionada, principalmente, pela redução na subunidade β da β -conglícinina (7S). Pode ser observado que o aumento do teor de proteína das isolinhas foi acompanhado, principalmente, pelo aumento na quantidade da proteína 11S. A soma dos conteúdos das duas proteínas (11S + 7S) aumentou quando houve aumento do teor de proteína, no entanto a quantificação individual mostra que este aumento ocorreu somente na proteína 11S,

havendo até mesmo uma pequena redução no teor da proteína 7S. Isso é de grande interesse, pois, a proteína 11S possui uma melhor qualidade nutricional e uma melhor digestibilidade em relação à proteína 7S. Estes resultados divergem em parte daqueles encontrados por YAKLICH (2001) que analisou diversas variedades e linhagens de soja de alto teor de proteína, mostrando que os teores de ambas as proteínas 7S e 11S aumentam nestas linhagens e variedades. Este autor também mostrou que os conteúdos dos polipeptídeos componentes das subunidades de cada uma das proteínas de reserva variam entre as diversas linhagens e variedades, e que foi detectado que algumas das linhagens de alto teor protéico apresentaram maiores acúmulos dos polipeptídeos α' , α e β componentes da proteína β -conglícinina, enquanto que todas as linhagens e variedades de alto teor protéico apresentaram maiores acúmulos dos polipeptídeos de carácter ácido e básico componentes da proteína glicínina quando comparadas com variedade padrão.

Tabela 1 – Teor de proteína, óleo carboidratos e cinzas das isolinhas 1 e 2 e da linhagem UFV-105, com base na matéria seca.

Amostra	Teor de proteína (%)	Teor de óleo (%)	Cinzas (%)	Teor de carboidratos totais (%)
UFV-105	40,68	20,78	4,93	33,61
Isolinha 1	47,78	16,81	5,40	30,01
Isolinha 2	46,56	16,71	5,66	31,07

Tabela 2 – Composição aminoacídica das farinhas desengorduradas da linhagem UFV 105 e das isolinhas 1 e 2. Os valores apresentados são médias de duas repetições e estão mostrados em porcentagem (%) do aminoácido na farinha de soja desengordurada e na proteína.

Aminoácido	UFV-105	%aa/Pt	Isolinha 1	%aa/Pt	Isolinha 2	%aa/Pt
Lisina	3,20	6,83	3,65	7,11	3,30	6,56
Histidina	1,33	2,84	1,54	2,99	1,38	2,74
Arginina	3,45	7,37	4,14	8,08	3,77	7,48
ác. Aspartico	6,16	13,14	6,29	12,26	6,44	12,79
Treonina	2,17	4,62	2,22	4,33	2,22	4,41
Serina	2,78	5,94	2,74	5,34	2,80	5,56
ác. Glutâmico	8,59	18,34	8,58	16,73	8,76	17,40
Prolina	2,40	5,11	2,79	5,44	2,65	5,25
Glicina	2,15	4,58	2,35	4,58	2,37	4,71
Alanina	2,13	4,54	2,26	4,41	2,29	4,54
1/2 Cistina	0,48	1,03	0,50	0,97	0,54	1,06
Valina	1,91	4,07	2,42	4,71	2,35	4,67
Metionina	0,62	1,31	0,67	1,31	0,66	1,30
Isoleucina	1,98	4,23	2,48	4,83	2,38	4,73
Leucina	3,63	7,75	4,24	8,27	4,00	7,95
Tirosina	1,57	3,34	1,80	3,51	1,76	3,49
Fenilalanina	2,32	4,94	2,63	5,13	2,70	5,35
Total	46,83	100,00	51,26	100,00	50,34	100,00

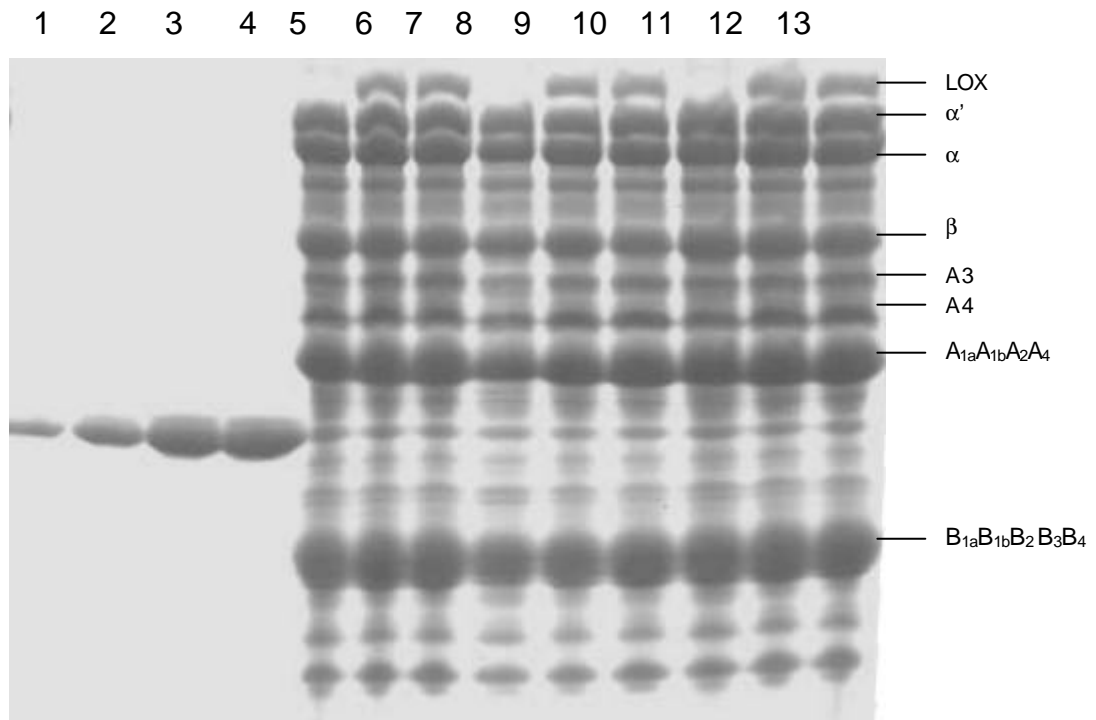


Figura 1 – Análise eletroforética SDS-PAGE em gel gradiente de 10 a 18% das proteínas de soja extraídas da linhagem UFV 105 e das isolinhas derivadas com alto teor de proteína. Canaletas 1, 2, 3 e 4 contêm 2, 4, 8 e 10 µg de proteína padrão anidrase carbônica. Canaletas de 5 a 13 contêm triplicatas de proteína total extraídas da linhagem UFV-105 e das isolinhas 1 e 2, respectivamente.

Tabela 3 – Quantificação das subunidades das proteínas de soja 7S e 11S por densitometria em gel gradiente de poliacrilamida 10-18%.

Subunidades	UFV-105	isolinha1	isolinha2
α'	0,25	0,22	0,22
α	0,40	0,37	0,37
β	0,33	0,26	0,26
A3	0,10	0,11	0,12
A4	0,13	0,19	0,20
ÁCIDAS	0,54	0,62	0,67
BÁSICAS	0,37	0,66	0,60

Tabela 4 – Quantificação das proteínas 7S e 11S por densitometria em gel gradiente de poliacrilamida de 10-18%.

Genótipos	7S	11S	11S/7S	11S+7S	%proteína	%(7S+11S)
UFV105	0,965	1,125	1,17	2,09	40,68	51,38
Isolinha-1	0,845	1,56	1,85	2,41	47,78	50,44
Isolinha-2	0,845	1,58	1,87	2,43	46,56	52,20

REFERÊNCIAS BIBLIOGRÁFICAS

- ASSOCIATION OF OFFICIAL ANALYTICAL CHEMISTS – AOAC. **Official methods of analysis**. Washington, D.C.: 1975. 1094p.
- ALONSO, N. HIRS, C.H.W. Automation of sample application in amino acid analysers. **Anal. Biochem.**, vol. 23, p.272-278, 1968.
- FONTES, E.P.B, MOREIRA, M.A., DAVIES, C.S., NIELSEN, N.C. Urea elicited changes in relative electrophoretic mobility of certain glycinin and β -conglycinin subunits. **Plant Physiol.** v.76, p.840-842, 1984.
- HARTWIG, E.E., KUO, T.M., KENTY, M.M. Seed protein and its relationship to soluble sugars in soybean. **Crop Sci.**, v.37, p.770-773, 1997.
- HELMS, T.C., ORF, J.H. Protein, oil and yield of soybean lines selected for increased protein. **Crop Sci.**, v.38, p.707-711, 1998.
- HILL, J.E., BREIDENBACH, R.W. Proteins of soybean seeds. II. Accumulation of the major protein components during seed development and maturation. **Plant physiol.**, v.53, p.747-751, 1974.
- IMSANDE, J. Selection of soybean mutants with increased concentrations of seed methionine and cysteine. **Crop Sci.** vol. 41, n. , p.510-515, 2001.
- INSTITUTO ADOLFO LUTZ. **Normas analíticas do Instituto Adolfo Lutz, Métodos químicos e físicos para análise de alimentos**. 3º ed., São Paulo, v.1, p.533, 1985.
- KITAMURA, K., KAIZUMA, N. Mutant strains with low level of subunits of 7S globulin in soybean (*Glycine max* Merr.) seed. **Japan. J. Breed.**, v.31, p.353-359, 1981.
- KITAMURA, K. Genetic improvement of nutritional and food processing quality in soybean. **JARQ**, v.29, p.1-8, 1995.
- MOORE, S.; SPACKMAN, D.H. & STEIN, W.H. Chromatography of amino acid on sulfonated polystyrene resins. **Anal. Chem.**, v.30, n.7, p.1185-1190, 1958.
- NIELSEN, N.C., JUNG, R., NAM, Y.W., BEAMAN, T.W., OLIVEIRA, L.O., BASSUNER, R. Synthesis and assembly of 11S globulins. **J. Plant Physiol.** v.145, p.641-647, 1995.

- OGAWA, T., TAYAMA, KITAMURA, K., KAIZUMA, N. Genetic improvement of seed storage proteins using three variant alleles of 7S globulin subunits in soybean (*Glycine max* L.). **Japan J. Breed.**, v.39, p.137-147, 1989.
- SPACKMAN, D.H.; STEIN, W.H. & MOORE, S. Automatic recording apparatus for use in the chromatography of amino acids. **Anal. Chem.**, v.30, n.7, p.1190-1206, 1963.
- WILCOX, J.R. Increasing seed protein in soybean with eight cycles of recurrent selection. **Crop Sci.** vol. 38, p.1536-1540, 1998.
- WILCOX, J.R., SHIBLES, R.M. Interrelationships among seeds quality attributes in soybean. **Crop Sci.** vol. 41, p.11-14, 2001.
- YAKLICH, R.W. β -Conglycinin and glycinin in high protein soybean seeds. **J. Agric. Food Chem.**, v.49, p.729-735, 2001.